UNIVERSITY OF OTTAWA

DOCTORAL THESIS

Essays in Public Economics and the Economics of Crime

Author:
Taylor WRIGHT

Supervisor:
Dr. Abel BRODEUR

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Economics
Faculty of Social Sciences
University of Ottawa

September 22, 2022

Declaration of Authorship

I, Taylor Wright, declare that this thesis titled, "Essays in Public Economics and the Economics of Crime" and the work presented in it are my own. Chapter one of this thesis is research done jointly with Dr. Abel Brodeur. My contribution is equal to his. The second chapter of this thesis is research done solely by myself. The third chapter of this thesis is work done jointly with Dr. Abel Brodeur and Dr. Nikolai Cook. My contribution is equal to theirs. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Taylor WRIGHT

Date: September 22, 2022

Acknowledgements

Gratitude goes first and foremost to my wife, Kari, for her patience, support, and sacrifice. Most importantly, she reminds me that I study people not data points. Without her, I would not be the person I am today.

I am indebted to my supervisor and coauthor, Abel Brodeur for being a constant source of inspiration and motivation. It was while taking one of Abel's classes that my love for research was piqued and I will forever be grateful for that. Abel's approach to supervision was an apprenticeship that gave me skills and confidence as a researcher while leaving room for my own development. What's more, his mentorship provided me with tacit knowledge of the profession that I am growing to appreciate more each day.

I am also thankful for my thesis committee members: Louis-Philippe Béland and Myra Mohnen who agreed to join only shortly after arriving in Ottawa and who have been invaluable for my development. As well as for my third committee member, Louis-Philippe Morin, who has always made time to give his insights and assistance. I am very grateful to Steven Lehrer for serving as my external examiner and imparting valuable comments.

Special thanks also go to Pierre Brochu, Catherine Deri Armstorng, Jason Garred, and Matt Webb who have provided excellent feedback and advice and have been constant supporters during my time in the program.

Lastly, I would like to thank Joshua Brault, Nikolai Cook, Lamis El Kattan, Fabien Forge, Joanne Haddad, and Florian Richard and my colleagues at the University of Ottawa and Carleton Unviersity for their feedback and support.

Contents

List of Figures

List of Tables

To my loving and patient wife, Kari

Chapter 1

Terrorism, Immigration and Asylum Approval

1.0.1 Abstract

Using the universe of individual asylum cases in the United States from 2000–2004 and a difference-in-differences research design, we test whether Sept. 11, 2001 decreased the likelihood that applicants from Muslim-majority countries were granted asylum. Our estimates suggest that the attacks resulted in a 3.2 percentage point decrease in the likelihood that applicants from Muslim-majority countries are granted asylum. The estimated effect is larger for applicants who share a country of origin with the Sept. 11, 2001 attackers. These effects do not differ across judge political affiliation. Our findings provide evidence that emotions affect the decisions of judges.

1.0.2 Thanks

We gratefully acknowledge comments from two anonymous referees, the co-editor of Journal of Economic Behavior & Organization Scott Adams, Jason Garred, Anthony Heyes, Paul Makdissi, Louis-Philippe Morin, Soodeh Saberian, and Matt Webb. Declarations of interest: none. Any remaining errors are our own.

1.1 Introduction

How do emotions such as anger, fear, or sadness influence peoples' decision making? An extensive body of research assesses the part extraneous information and emotions play in how people make evaluations and judgments. However, relatively little analysis has investigated the effect of catastrophic events such as Sept. 11, 2001 and their resulting emotions on decision making.

One of the founding principles of the United States is equal and fair treatment under the law. The influence of emotions and extraneous information on judges' decision making is a deviation from this principle and a miscarriage of justice. Moving from the macro to the micro, the individuals applying for asylum (asylees) are doing so in order to flee persecution from their country of origin on the grounds of race, religion, nationality, belonging to a particular social group, or political opinion. Asylees who are deported unjustly face poverty, violence, torture, and/or death (see for example Andrews et al., 2019, Brodzinsky, 2015, Bulman, 2018, or Stillman, 2018). Clearly, these are high stakes decisions that require the evaluation and removal of underlying judicial prejudice. We believe that documenting the magnitude of the effect of terrorist attacks on asylum approval is therefore meaningful and important.

Islamic terrorism may affect judges' decision making through several different mechanisms. Emotions have been shown to not only color the immediate and related judgments and evaluations, but also those that are unrelated and occur much later (Lerner et al., 2015). According to this mechanism, the Sept. 11, 2001 attacks represent a negative event and as a consequence we might expect to see lower asylum grant rates. If cases of asylees from Muslim-majority countries create emotional cues related to the attacks, we might expect the decline in asylum grant rates to be even larger for asylees from those countries.[1]

Our empirical strategy investigates how Sept. 11, 2001 (and the March 11, 2004 attacks that occurred in Madrid, Spain) changed asylum grant rates for asylees from Muslim-majority countries in American immigration courts. Using the universe of asylum cases in the United States from January 2000–September 2004, we employ a difference-in-difference research design with applicant and temperature controls alongside judge, city, month, and city-month fixed effects. We address potential selection issues by exploiting random judge assignment; the presence of lengthy, case backlogs; and creating windows around the attack date.

Our estimates suggest that asylees from Muslim-majority countries are about 3.2 percentage points less likely to receive asylum than those not from Muslim-majority countries. We also provide evidence that asylees from attack associated countries are between 6–10 percentage points less likely to receive asylum than those not from countries associated with Sept. 11, 2001. We find asylees from Muslim-majority countries are also about 2.2 percentage points less likely to receive asylum after the Madrid attacks. Our results also suggest that Democrat appointed judges do not respond to the attacks in a different way than do Republican appointed judges.

This paper is related to the literature on emotional shocks and decision making, extraneous information and judicial bias, and the consequences of terrorism. Several studies examine the effects of negative emotional shocks on various outcomes. Beland and Brent, 2018 find that emotional shocks from extreme traffic in Los Angeles increases domestic violence while Card and Dahl, 2011) find that unexpected local football losses increase domestic violence in the United States. Eren and Mocan, 2018 find that unexpected college football losses by a judge's alma mater result in harsher juvenile sentencing in Louisiana.

Our paper also relates a growing literature examining the impact of extraneous information on the decisions of judges (e.g. Danziger, Levav, and Avnaim-Pesso, 2011). Heyes and Saberian, 2019 examine the impact of pollution and temperature on immigration judge asylum decisions, finding that extreme temperature results in lower likelihood of asylum. Philippe and Ouss, 2018 explore the impact of crime on jury decision making in France and conclude that new coverage of unrelated crimes increases sentence length. Shayo and Zussman, 2011 find that Arab and Jewish small claims judges are biased in favor of their own groups, and that this bias is strongly related to the intensity of nearby terrorism in the previous year. Our work is also very closely related to McConnell and Rasul, 2019, who examine the spillover effects of the Sept. 11, 2001 attacks on differential sentencing of hispanic and black defendants in the U.S. finding that hispanic defendants experienced worse outcomes as a consequence.[2]

[1]Other possible channels include judges updating their perceptions of the frequency of Islamic terrorism or terror-related changes in local economic conditions. See Section 1.3 for more details.

[2]Also of note is work examining racial bias in the judicial system: Alesina and La Ferrara, 2014 find evidence of bias in U.S. capital sentencing against minority defendants; Anwar, Bayer, and Hjalmarsson, 2012 find bias in Florida conviction rates of black defendants when there were no black jury

Our work contributes to these existing behavioral literatures in two key ways: first, we examine the effect of a single catastrophic Islamic terrorist attack (and its resulting emotional shock) at a time in the U.S. when there had been essentially no Islamic terror attacks committed;[3] second, we are examining the decisions of immigration judges rather than those of small claims judges or juries where the degree of judge discretion and stakes of the decisions are quite different.

Previous empirical work has focused on examining the effect of terrorism on many outcomes which, although closely related, differ from ours. We contribute here by documenting the effects of terrorism on asylum approval for those from Muslim-majority countries, and judge decisions. Dreher, Gassebner, and Schaudt, 2017 estimate the impact of changes in immigrant stocks on the probability of terror attacks in the host country, finding no evidence that terrorism is imported from Muslim countries or other countries with significant terrorism.[4] Cornelissen and Jirjahn, 2012, Dávila and Mora, 2005, Kaushal, Kaestner, and Reimers, 2007 document the impacts of Sept. 11, 2001 on labor market outcomes of Arabs and Muslims.[5] Our work also ties in with studies examining the effect of terrorism on changes in ethnic attitudes (Ratcliffe and Hinke Kessler Scholder, 2015), self-identification among Arab and Islamic Americans (Mason and Matella, 2014) and integration and assimilation (Bisin et al., 2008, Elsayed and De Grip, 2018, Gould and Klor, 2016).

Last, our work complements studies documenting the consequences of transnational terrorism on fear, uncertainty, and behavioral responses to those emotions. We mostly relate to the contributions of Becker and Rubinstein, 2011 who argue that terrorism may lead to intense fear of future dangers and Brodeur, 2018 who provides empirical evidence that terror attacks in the U.S. decrease consumer confidence.

The paper is structured as follows. Section 1.2 provides background for the the entry of asylum seekers into the U.S. and the role judges play in the asylum process. In Section 1.3, we provide a conceptual framework and review the literature on behavioral biases. Section 1.4 describes the data used and provides summary statistics. Section 1.5 outlines our empirical strategy and model specifications. Section 1.6 presents our results. The last section concludes.

1.2 Asylum and Immigration Judges

In this section, we first briefly describe how immigration judges are hired, by whom, and what their qualifications and backgrounds look like. We then provide an overview of the asylum process in the United States.

1.2.1 Immigration Judges and the Executive Office for Immigration Review

Asylum adjudication is carried out by the Executive Office for Immigration Review (EOIR) which was created in 1983 after a restructuring of the Department of Justice

members; and Cohen and Yang, 2019 find that U.S. Republican-appointed judges deliver harsher sentences to black defendants and the difference between white and black defendants grows with judge discretion.

[3]Therefore the social division generated by these attacks is significantly less ingrained than in an ongoing conflict context.

[4]Avdan, 2014 finds that terrorism in Europe results in migration restriction for countries who experience attacks, but does not result in the erosion of the humanitarian principles backstopping asylum recognition.

[5]See Abadie and Gardeazabal, 2003, Blomberg, Hess, and Orphanides, 2004, and Crain and Crain, 2006 for the macroeconomic consequences of terrorism.

(DOJ). Its creation separated asylum adjudication from the enforcement of immigration laws (formerly the Immigration and Naturalization Service (INS) which is now part of the Department of Homeland Security (DHS)). The EOIR is tasked with carrying out immigration court hearings, administrative hearings, and appellate reviews and does so with delegated authority of the Attorney General. The Director of the EOIR reports directly to the Deputy Attorney General (Department of Justice, 2018a).

Formally, the Attorney General makes immigration judge appointments; however, the hiring process is conducted by the EOIR. The requirements for becoming an immigration judge include being a U.S. citizen or national in possession of a law degree with at least 7 years of post bar legal experience who is authorized to practice law as an attorney in the U.S. Applicants are also required to submit a writing sample to demonstrate their ability to author legal documents (Department of Justice, 2018b).

We manually collected judge characteristics for all the judges in our dataset (see Section 1.4 for more details) and use those with full enumeration as controls. Unfortunately, the data is incomplete as we are missing complete biographical information for about 10 percent of our judges. Of the judges for which we have complete information, about 50 percent had previously worked for the INS in some capacity, often as trial attorneys. Around 45 percent of judges had previously worked in legal aid, nearly the same proportion as having worked for a firm or in private practice. These background characteristics are not mutually exclusive and it is common for judges to, for example, have worked for the INS and also have worked for a firm or legal aid organization.

Immigration courts are often not staffed with law clerks to provide additional research assistance or bailiffs (unless the hearing are taking place at a detention center) and judges are required to manage their own recording of proceedings (National Association of Immigration Judges, 2011). These resource issues are compounded by a heavy workload—making hundreds of decisions each year and hearing several times that many—and hundreds of thousands of backlogged cases.[6] Perhaps as a result of these conditions, the EOIR and DOJ are not able to fill vacancies in a timely manner and there have been concerns with the temperament, quality, and performance of judges from Federal courts and the Attorney General (Cable News Network (CNN), 2006; Liptak, 2005; National Association of Immigration Judges, 2011).

1.2.2 The Asylum Process

Asylees apply for refugee status from within the United States, either in response to being deported or proactively, to escape persecution in their own country. Their cases are heard by randomly assigned immigration judges who are institutionally independent of immigration enforcement, typically in the court closest to their physical location at the time of the application (Miller, Keith, and Holmes, 2015; Refugee, Asylum, and International Operations Directorate, 2016). Hearings are adversarial and asylees must provide their own legal representation (Ramji-Nogales, Schoenholtz, and Schrag, 2007). In our empirical analysis, we control for whether applications are affirmative or defensive, whether or not the asylee has legal representation, and whether or not the asylee was detained. We also use judge fixed effects (or in

[6]There are currently over 800,000 cases that are pending but undecided. In our sample, there were approximately 150,000-200,00 pending cases (TRAC, 2018).

the case of our DDD analysis control for judge gender, experience, and the party in control of the White House when they were appointed).

1.3 Conceptual Framework

Immigration judges are required to determine whether or not the fear of persecution is "well-founded". The fundamental directive immigration judges are given is to evaluate the eligibility and assess the likelihood of an asylee's persecution upon return to their country of origin.[7] Nonetheless, immigration judges have significant discretion in adjudicating cases—the definitions of terms like "persecution" used in establishing eligibility for asylum are vague or left up to individual judges as is the estimation of probabilities of persecution.

There are thus several channels through which Islamic terror attacks could influence immigration judges' decision making. We first focus our attention on unconscious or behavioral mechanisms. We then discuss other mechanisms through which Islamic terrorism could affect judges' decisions.

1.3.1 Behavioral Biases

The institutional setting and context in which immigration judges operate may make them more likely to bring their policy preferences and personal biases to bear on the cases they examine, using them to filter case facts, regardless of legal merit.

There is substantial evidence from economics, psychology, and neuroscience that emotions influence decision making (see Lerner et al., 2015 for a an overview). This is relevant in our context since negative emotions may lead to pessimistic evaluations (Johnson and Tversky, 1983).[8] This is complemented by findings that emotions persist, continuing to influence decisions that are unrelated to the emotion (e.g., Han, Lerner, and Keltner, 2007). Again, this is important for our context as, if emotions persist, those pessimistic evaluations can occur in a context entirely divorced from that which generated the emotion. There is a burgeoning empirical literature documenting this concept. For example, Lerner et al., 2003 conduct a field experiment in the aftermath of Sept. 11, 2001 where participants are treated with news articles inducing fear or anger. They find that those affected by fear viewed the world with greater risk and those with anger with less risk. Important for our research, those affected by anger supported more harsh treatment of suspected terrorists than those affected by fear.

Two other relevant studies are Eren and Mocan, 2018 and Philippe and Ouss, 2018. Philippe and Ouss, 2018 examine the effect of news coverage of unrelated crime on juror sentences and find an increase in sentence duration. They also explore whether or not this is something specifically related to crime or whether or not crime is another form of bad news, finding that other forms of bad news that may affect mood do not influence juror decisions. Eren and Mocan, 2018 explore the effect of

[7] The lack of research resources and time constraints they face may make them reliant on existing reports produced by the State Department (Miller, Keith, and Holmes, 2015).

[8] Danziger, Levav, and Avnaim-Pesso, 2011 find that the order in which parole judges review cases and the timing of food breaks influences judge decision making, though whether the mechanism is the break, food, or possible change in mood is unclear.

unexpected football team losses on juvenile court judges in Louisiana and find that emotional shocks result in harsher sentencing of black juveniles.[9]

To the extent that judge characteristics determine which emotion was felt by the attack itself, the attacks may influence judge behavior in different ways. These potentially heterogenous emotional responses combined with the importance of immigration judge policy preferences are key reasons for our collection and use of experience and appointing political party as controls and factors that may generate heterogenous treatment effects.

1.3.2 Other Mechanisms

Terrorism and the associated media coverage could shift the perceptions of the frequency of Islamic terrorism. This is a cognitive effect known as the "availability heuristic", which is a mental shortcut in which a person's perceived frequency of an event is influenced by how easily they can call to mind an instance of that event (Tversky and Kahneman, 1973). While the change in perceived frequency is an unconscious decision, the decision to rely on "gut feelings" and avoid challenging or updating those perceptions is conscious. This effect could also interact with the country of origin, as Islamic terrorism could alter the perceived frequency of individuals from Muslim countries being terrorists. These two consequences of Islamic terrorism offer an alternative, conscious mechanism that explains changes in judges' decision making, specifically for asylees from Muslim-majority countries.

In Section 1.1 we referenced the body of existing research documenting the consequences of terrorism on assimilation, integration, and labor markets. If immigration judges are influenced by local conditions then this existing research suggests a mechanism other than emotion through which judges' decision making could be altered. Miller, Keith, and Holmes, 2015 suggest that judges can be influenced by local conditions, drawing on other research indicating that judges' decisions in other contexts correlate strongly with local attitudes; that opposition to immigration is tied to local labor market conditions; and that immigrants are a fiscal burden on their local governments.

1.4 Data

Our data on asylum decisions is administrative, case-level data containing the universe of asylum cases in the United States occurring between January 2000 and September 2004. Our data comes from asylumlaw.org, a now defunct website that was a collaborative effort by international agencies to assist asylum seekers in many countries, including the United States. The website contained information and resources including those related to legal and human rights; and immigration and asylum experts.[10] In total, we have 269,270 asylum decisions made by 262 immigration judges across courts based in 43 cities.

The dataset contains information about the date and location of hearing, the judge responsible for the case, whether or not an applicant had legal representation, the nationality of the applicant, whether or not the application was in response to deportation proceedings. We merge this data with information about the share of

[9] Card and Dahl, 2011 find that unexpected football team losses cause an increase in domestic violence mediated by a change in mood. They describe a 'gain-loss' utility framework where sport team losses result in fluctuations around a rational reference point.

[10] Data retrieved from Heyes and Saberian, 2019.

the population that is Muslim in the applicant's country of origin.[11] Additionally, we are able to determine whether or not an applicant was detained in an institution during their hearing. The dataset contains information not only on cases where some form of asylum is denied or approved, but also cases that are withdrawn or dismissed. These applicants are less likely to have legal representation (69%) and less likely to be filing defensively (21%). In our analysis, we include dismissed and withdrawn cases but our results are robust to their exclusion (see Section 1.6).

We collected biographical information about each judge in the sample. Our primary data source for collecting judges' characteristics is the judge reports from the Transactional Records Access Clearinghouse (TRAC) which is a data collection, research, and data warehouse based out of Syracuse University. Their website contains reports for judges each year in which the judge decides at least 100 cases. The judge information was collected from "a variety of official sources including press releases, testimony, other biographical information released by the Department of Justice, and responses received to specific TRAC inquiries" (TRAC, 2008). Unfortunately, the 100 decision per year threshold in the TRAC data leaves 30 of the 262 judges missing. We used internet searches, relying on DOJ and other government documents or newspaper articles, to compile information about these judges. We were able to obtain gender for all judges and year of appointment for all but four. Unfortunately, information on the backgrounds of these judges was not as readily available and as such, we do not use this information in our analysis.

We collected information about judge gender, year of appointment, and backgrounds prior to appointment. We used the year of appointment to construct variables indicating the amount of experience a judge has at the time of Sept. 11, 2001. Last, we construct a dummy variable for which political party held the Presidency at the time of the judge's appointment. This variable serves as a measure of a particular judge's ideology as judges are hired by the executive branch of the U.S. government.

There is a large degree of variation in asylum approval rates between courts. For instance, in 2001, the court in San Francisco, California approved 65 percent of cases while the court in Philadelphia, Pennsylvania approved 29 percent of cases. There is also a large degree of variation in approval rates within courts: approval rates in 2001 ranged from 6 percent to 93 percent in New York, New York for judges who adjudicated at least 200 cases. Importantly for our analysis, cases are randomly assigned to immigration judges within a court. In the absence of random case assignment our identification strategy would be threatened by sorting of Muslim cases to judges less (or more) likely to grant asylum.

Table 1.1 provides summary statistics across our sample on asylum outcomes and case characteristics for the U.S. as a whole. Each of the variables contained in Table 1.1 are binary variables and thus the means presented represent the share of cases for which the statement is true. The mean grant rates for asylees is 19.9% with a standard deviation of 39.9. For asylees from Muslim-majority countries the mean grant rate is 28.3% with a standard deviation of 45.1. Over 80 percent of asylees have a legal representative and 8.2 percent were detained. Just under 30 percent of applications are defensive (in response to removal proceedings) and just over 70 percent are affirmative (no removal proceedings being pursued).[12] About 18 percent

[11] Data about the share of the population that is Muslim in each country is drawn from 2010 estimates by the Pew Research Center's Forum on Religion and Public Life and we define a country as Muslim-majority if greater than 50% of their population is Muslim. Note that our definition of Muslim-majority countries would not change if we were using 1990 instead of the 2010 estimates.

[12] Removal proceedings may occur as a result of criminal convictions, determination of illegal entry into the United States, or possession of insufficient documentation at border crossings.

of asylees have a Muslim-majority country of origin and 1.3 percent share a country of origin with a perpetrator of the Sept. 11, 2001 attacks.

Around 70 percent of applications were heard by judges appointed by a Democrat. Applications were heard by female judges around 36 percent of the time. Just over 43 percent of applications were heard by a judge who had between 0 and 5 years of experience. Judges between 6 and 10 years of experience heard around 21 percent of applications while those with over 10 years of experience handled the remaining 34 percent or so.

1.5 Identification Strategy

The objective is to estimate the impact of terrorism on judges' decisions when granting asylum. We rely on two empirical models.

Difference-in-Differences – In our main empirical analysis, we implement a difference-in-differences (DD) research design by comparing the asylum grant rates of asylees from Muslim-majority countries and asylees not from Muslim-majority countries, before and after Sept. 11, 2001. In our main specification, we estimate

$$y_{itmjc} = \alpha + \beta Muslim_i + \delta Post911_t + \gamma Muslim_i \times Post911_t + \\ X_i'\psi + \theta temperature_{ct} + \lambda_j + \tau_c + \omega_m + \rho_{cm} + \varepsilon_{itmjc} \quad (1)$$

where y_{itmjc} is a binary variable that equals one if a judge j grants asylum for case i on date t in month m in city c. $Muslim_i$ is a binary variable that equals one if the asylee is from a Muslim-majority country and zero otherwise. $Post911_t$ is a dummy that equals one if a case was heard after the attacks. Our coefficient of interest in this model is γ. It shows the effect of DD interaction term, $Muslim_i \times Post911_t$. X_i' is a vector of case specific variables including whether or not the individual is detained at the time of the hearing, has legal representation, and is filling affirmatively or defensively. $temperature_{ct}$ measures the mean temperature from 6am to 4pm in city c on date t and is included following Heyes and Saberian, 2019. λ_j, τ_c, ω_m, and ρ_{cm} are judge, city, month, and city-month fixed effects, respectively.

We also estimate Equation (1) replacing $Muslim_i$ with a binary variable $Associated_i$ that equals one if the asylee is from a country associated with the Sept. 11, 2001 attacks and zero otherwise. Associated countries are defined as the countries of origin of the 19 attackers: one from Egypt, one from Lebanon, two from United Arab Emirates, and 15 from Saudi Arabia. This alternative specification tests whether the effect of Sept. 11, 2001 is larger/smaller for this subgroup of asylees. We hypothesize that the effect is larger since the behavioral biases discussed in Section 1.3 could be magnified for asylees from these countries. Again, the coefficient of interest is the DD interaction term, this time between $Associated_i$ and $Post911_t$. We exclude Muslim-majority countries other than those associated with the attack, making the comparison between asylees from attack-associated countries and those from non-Muslim-majority countries.

Additionally, we estimate Equation (1) replacing $Post911_t$ with a binary variable $Post311_t$ that equals 1 if the case is heard after the March 11, 2004 terror attacks conducted in Madrid, Spain. The coefficient of interest in these regressions is the interaction $Muslim_i \times Post311_t$.

We consider the treatment as being applied at the country level with Muslim-majority countries as being treated and non-Muslim-majority countries as untreated. We cluster our standard errors at the country of origin level because it seems likely that the unobservables of asylees will be correlated within countries (e.g. spoken languages, or religion). We have 192 clusters in total.

Our identification assumption is that in the absence of the Sept. 11, 2001 attacks, the change in the grant rates of Muslim-majority countries and other countries would not be different, conditional on controls. We visually demonstrate that this parallel trends assumption holds for the pre-treatment period. Figure 1.1 plots the monthly average asylum grant rate over our entire sample (January 2000 to September 2004) for asylees from Muslim-majority countries and asylees not from Muslim-majority countries. The vertical line represents the Sept. 11, 2001 attacks. Figure 1.2 plots the analog for asylees from countries associated with the attacks. In Figure 1.1, we can see that prior to the attacks, the grant rates for both groups were trending upwards. Asylees from Muslim-majority countries were much more likely to be granted asylum than those from non-Muslim countries and it appears as though there is a slight convergence in grant rates in the post-attack period. In the year or so prior to the attacks, the monthly average for asylees from Muslim-majority countries was nearly as likely to be above 30% as below it and was above 30% eight times. In the just over 2 years after the attacks, however, only six months broke 30% and several months fell at or below below 25%. For those from non-Muslim countries it appears there is a leveling off of the growth in grant rates after the attacks with the average asylum grant rate in most months coming in between 15%–20%.

Switching to Figure 1.2, we can see that asylees from countries associated with the attacks were much more likely to be granted asylum both prior to the attacks and after the attacks. However, just as in Figure 1.1 the gap appears to shrink after the attacks. For asylees from associated countries, prior to the attacks, the average asylum grant rate for several months broke 60% and only once was below 30%. After the attacks, several months fell below 30% and no months broke 70%. As with Figure 1.1, there appears to be an upward trend in the average monthly asylum grant rate prior to the attacks. Afterward, the growth rate levels out for asylees from non-associated countries and appears to be slightly negative for those from associated countries. The results for asylees from associated countries are noisier than those for non-associated countries, possibly because of the smaller number of observations for asylees from associated countries.

In both figures, the trends for asylees from Muslim-majority countries and associated countries appear to track the trends for asylees not from Muslim-majority or associated countries very closely. More formally, Appendix Table A1 shows the estimation of Equation 1 with the inclusion of a linear time trend interacted with a dummy variable indicating if the asylee is from a Muslim-majority country. The full suite of controls are also included and standard errors are clustered at the country of origin. The results indicate that there is no significant difference in the pre-treatment trends for asylum approval rating between asylees from Muslim-majority countries and those not from Muslim-majority countries. Together, these provide evidence that our choice of control groups represent appropriate counterfactuals prior to the Sept. 11, 2001 attacks and would continue to do so afterward. We delay our discussion of placebo treatments to Section 1.6.

Given the large degree of heterogeneity in judge grant rates, it is crucial for our identification that there is no sorting of cases for asylees from Muslim-majority countries to judges that are particularly lenient or severe. Appendix Table A2 presents the results from a regression of a judge stringency measure (created by calculating the leave-one-out mean grant rates for each judge) on a dummy variable that equals 1 if an asylee is from a Muslim-majority country and 0 otherwise, including court fixed-effects. If cases are randomly assigned, the point estimate should be close to zero and that is exactly what we find.

There may be concerns that Sept. 11, 2001 and March 11, 2004 resulted in a change in the composition of asylum applicants. This would be an issue if the change is more pronounced for Muslim applicants, possibly in anticipation of potential backlash. This is not an issue for our main specification since we restrict the sample only to cases occurring six months before and after Sept. 11, 2001. The Immigration and Nationality Act requires that a decision be made on applications within 180 days of filing, though in practice backlogs are much longer than this. According to the Transactional Records Access Clearinghouse, a data gathering, distribution, and research organization housed at Syracuse University, the average wait time for the entire U.S. across all nationalities ranged from a low of 380 days in 2000 to a high of 422 days in 2003 (TRAC, 2018), while the average processing time ranged from 234 days in 2001 to 283 days in 2004 (TRAC, 2018). As a consequence, the hearings in our sample will correspond to applications filed before Sept. 11, 2001 and March 11, 2004 thus avoiding any potential bias from compositional changes in the applicant pool. We also note that due to data limitations, we are only able to conduct 3 month and 6 month windows around March 11, 2004.[13]

Additionally, we include Appendix Table A3 which regresses whether an asylee is from a Muslim-majority country on a binary variable that equals 1 if the case is heard after Sept. 11, 2001. The estimated coefficient is statistically insignificant, meaning that hearings occurring after the attacks are no more likely to be for asylees from Muslim-majority countries than are those prior to the attacks.

Triple Differences – We supplement our main analysis by investigating how the effect estimated in Equation 1 varies across immigration judges appointed during a Republican or Democrat presidency, employing a triple-differences (DDD) research design [14]. We estimate

$$\begin{aligned} y_{itmjc} = \alpha + \beta Muslim_i + \delta Post911_t + \zeta Muslim_i \times JudgeCharj + \\ \xi Post911_t \times Democrat_j + \kappa Muslim_i \times Post911_t \times Democrat_j + \\ \gamma Muslim_i \times Post911_t + X_i'\psi + V_j'\phi + \theta temperature_{ct} + \tau_c + \omega_m + \rho_{cm} + \epsilon_{itmjc} \end{aligned} \quad (2)$$

where V_j' is a vector of judge specific variables including gender, and dummies for between 0–5 years, 6–10 years and more than 10 years of experience as of Sept. 11, 2001. $Democrat_j$ is a binary variable that equals one if the judge was appointed during a Democrat presidency and zero if appointed during a Republican presidency. For example, we compare the difference in asylum grant rates for asylees from Muslim-majority countries (treated) to asylum grant rates for asylees from other countries (control), across Democrat (treated) and Republican (control) judges, before (control) and after Sept. 11, 2001 (treated). Our coefficient of interest in this model is κ, which shows the effect of the DDD interaction term, $Muslim_i \times Post911_t \times Democrat_j$. We do not estimate Equation (2) for asylees from countries

[13] We are seven days short of fully having 6 month windows around March 11, 2004.

[14] For other examples of triple-differences see, for example, Brodeur and Connolly, 2013, Jayachandran and Lleras-Muney, 2009, or Carr and Doleac, 2018

associated with the attacks due to sample size concerns. All other components are defined as in Equation (1).

1.6 Results

In this section, we first estimate the effects of Sept. 11, 2001 and March 11, 2004 on granting rates for applicants from Muslim-majority countries and countries associated with the attacks. We then estimate how the effect of Sept. 11, 2001 for Muslim-majority countries and associated countries varies by a judge's political affiliation. We conclude this section with robustness checks.

1.6.1 Difference-in-Differences

Table 1.2 presents OLS estimates of Equation (1) for asylees from a Muslim-majority country whose case was heard within 180 days before or after Sept. 11, 2001. What clearly emerges is that Sept. 11, 2001 is associated with a large decrease in the likelihood applicants from Muslim-majority countries are granted asylum. As noted in Section 1.5, we restrict our sample around the attacks on Sept. 11, 2001 in order to help address concerns about changes in the composition of applicants, perhaps reflecting strategic behavior in response to the attacks. Consequently, the time period is March 15, 2001 to March 9, 2002 (180 days before and after). The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The sample size is 48,340 observations (i.e., cases). As mentioned in Section 1.5, we report standard errors clustered by asylee country of origin in parentheses. Appendix Table A4 presents the estimated coefficients for control variables.

Column 1 presents the simple differences of an applicant being from a Muslim-majority country and having their case heard after Sept. 11, 2001. We show that prior to the attacks, applicants from Muslim-majority countries, in comparison to other applicants, were significantly more likely to receive asylum (coeff. of 0.112). The dummy $Post911_t$, which indicates whether the case was heard after Sept. 11, 2001, is negative and statistically significant at the 1% level. The estimate suggests that the attacks decreased the likelihood to receive asylum by about 2 percentage points (pp).

Column 2 adds our coefficient of interest, γ, in the third row. We find that the Sept. 11, 2001 attacks resulted in a 3.8 percentage point decrease in the likelihood that applicants from Muslim-majority countries are granted asylum. Column 3 adds applicant and temperature controls along with judge and city fixed effects, while columns 4 and 5 add month fixed effects and city-month fixed effects, respectively. Applicant controls include detention status, application type (affirmative or defensive), and legal representation, while temperature is the mean temperature in the city of the hearing from 6am to 4pm on the day of the hearing.[15] Our estimates are remarkably stable (3.2–3.8pp throughout) and are statistically significant at the 1% level across all specifications.

Our estimates suggest that Sept. 11, 2001 caused a 11.3%–13.4% drop in the likelihood that asylees from Muslim-majority countries receive asylum. In our sample, the mean grant rant for asylees from Muslim-majority countries is 28.3% (standard deviation 45.1). In order to gain insight into the magnitude of our estimates, we

[15] As discussed in Section 1.5, we control for these additional variables to ensure that our estimates are not picking up compositional changes in the pool of applicants in response to the events of September 11, 2001.

compare them to other studies. For instance, Heyes and Saberian, 2019 find that a one standard deviation increase in case-day temperature reduces asylum grant rates by 8.56%. Eren and Mocan, 2018 find that surprise loses of a football team from that judge's former college result in juvenile sentences that are 6% longer. Philippe and Ouss, 2018 find that french jury trials occurring after news on crime increases sentence length by 2.4%.

Table 1.3 presents the results of Equation (1) for asylees from a country associated with the Sept. 11, 2001 attacks whose case was heard within 180 days before or after the attacks. The structure of the table is the same as in Table 1.2. Once again, the attacks result in a large decrease in the likelihood of granted asylum and in fact the estimates here are much larger. As before, column 1 presents the simple differences of an applicant having their case heard after the attacks or of being from an associated country. We find that those from an associated country are significantly more likely to receive asylum. We also find a very similar estimate of the effect of an applicant's case being heard after Sept. 11, 2001.

The estimates in columns 2–5 suggest that the Sept. 11, 2001 attacks caused a significant decline in the likelihood of applicants from associated countries to receive asylum. The estimates range from 6pp to 9.9pp and are statistically significant at the 5% level in our preferred specification. It is important to emphasize that because there are relatively few asylees from attack associated countries, our estimates are relatively imprecise.

Table 1.4 contains the OLS estimates of Equation (1) for asylees from a Muslim-majority country whose case was heard within 180 days of the March 11, 2004 attacks in Madrid. The structure of the table is the same as Tables 1.2 and 1.3. Unlike our estimates for the Sept. 11, 2001 attacks, our estimates in column 1 of the effects of applicants being from Muslim-majority countries or for having a case heard after the attacks are not statistically significant at conventional levels. Columns 2–5 again suggest that we find an effect of the Madrid attacks of between 2.2pp and 2.9pp. These estimated effects are very similar in magnitude to those of the Sept. 11 attacks.[16] It may be surprising that the magnitude of the effect of the Madrid attacks on immigration judges is so similar to that of the Sept. 11 attacks which took place on American soil. One possibility is that immigration judges in America became more sensitive to terrorism related shocks, perhaps in part due to the political climate and perception of the risks and threat of terrorism post Sept. 11, 2001.

1.6.2 Robustness Checks and Placebo Test

Table 1.5 explores the sensitivity of our findings to alternative choices of pre- and post-Sept. 11, 2001 periods. The sample window varies from 3 months (90 days) to 12 months (365 days) before and after Sept. 11, 2001. Columns 1–4 correspond to applicants from Muslim-majority countries and columns 5–8 correspond to applicants from countries associated with the attacks. Each column includes all applicant and temperature controls along with judge, city, month, and city-month fixed effects. Column 2 is the same as column 5 of Table 1.2 and column 6 is a repetition of column 5 in Table 1.3. The results suggest two main things: (1) the estimated effect size for applicants from Muslim-majority countries is in the same ballpark across

[16]We also estimate this specification for the Madrid attacks with the treatment definition being changed to asylees sharing a country with the attackers (Morocco and Tunisia). Our estimated coefficients for the interaction term are negative but not statistically significant at conventional levels. However, the Madrid attacks were perpetrated by individuals linked to only two countries and thus we are quite underpowered for this analysis.

these different windows; and (2) the sample size issues mentioned in the previous section have a dramatic effect on the precision of the estimates for attack associated countries across different windows.

Additionally, Appendix Tables A5 and A6 present Logit and Probit results for Equation (1) for asylees from Muslim-majority countries whose case was heard within 180 days before or after Sept. 11, 2001. These tables are replications of Table 1.2. The estimates provided are marginal effects and are very similar to those found in Table 1.2. We find an effect of a magnitude around 2.2pp–2.8pp that is statistically significant at the 5% level and that is robust to adding or removing controls and fixed effects. These tables present a 6 month (180 day) window, but the results are robust across all windows.

Appendix Tables A7, A8, and A9 contain the results of placebo tests using Equation (1) for asylees from Muslim-majority countries and associated countries whose case was heard within 180 days before or after Sept. 11, **2000**, **2002**, and **2003**, respectively. Columns 1–3 correspond to applicants from Muslim-majority countries and columns 4–6 correspond to applicants from countries associated with the attacks. Because there were no attacks on Sept. 11, 2000, 2002, or 2003 the DD interaction terms should be zero. In fact, we see that these estimated effects are not significantly different from zero and are unaffected by including or excluding applicant and temperature controls and judge, month, and city fixed effects. While this table presents a 6 month (180 day) window, the results are robust across all windows. These placebo tests combined with our formal pre-trend test for divergence in Table A1 and visual inspection of the trends in Figures 1.1 and 1.2 suggest that non-Muslim asylees are an appropriate control group.

Appendix Table A10 produces the results of a placebo test using Equation (1) for asylees from Muslim-majority countries whose case was heard within 180 days of March 11, **2003**. Again, we expect to see that the DD interaction term should be 0 as there is no attack on March 11, 2003. We do in fact see that the estimated coefficient is not statistically different from 0 and robust to the inclusion or exclusion of controls and fixed effects.

Appendix Table A11 presents the same results as Table 1.2 but excludes withdrawn and dismissed cases. The estimated coefficient of interest is in the same direction and very close in magnitude to that presented in Table 1.2 and is statistically significant at the 5% level. Our other estimates using dismissed and withdrawn cases are similarly unchanged from those excluding them.

Appendix Table A12 provides estimates for Equation (1) raising the threshold for determining that a country is a Muslim-majority country from 50% to 75%. The estimates have the same direction and similar magnitude as those in Table 1.2 and the estimate in our preferred specification (column 5) is statistically significant at the 5% level.

1.6.3 Triple-Differences

Table 1.6 presents the results of Equation (2) for asylees from Muslim-majority countries. As in the previous DD analysis, the first two differences remain before and after Sept. 11, 2001 and whether or not the applicant is from a Muslim-majority country. The third difference is whether or not the case is heard by a judge who was appointed during a Democrat presidency. It is unclear whether there is a differential in the effect of Sept. 11, 2001 by the appointing political party of the judge given that approval of the Patriot Act was heavily bipartisan. Furthermore, the majority of Democratic Senators also approved the Authorization for Use of Military Force Against Iraq Resolution.

As in Table 1.5, we allow the sample window to vary from 3 months (90 days) to 12 months (365 days) before and after Sept. 11, 2001. All estimates include applicant, judge and temperature controls as well as city, month, and city-month fixed effects with standard errors clustered at the asylee's country of origin. Our results suggest that the differential effect of Sept. 11, 2001 on Muslim-majority country asylees does not vary by the appointing political party of the judge. Estimates are not statistically significant across any of the windows around the attacks and the point estimates of the coefficients are quite small. These results are in contrast with recent work by Cohen and Yang, 2019 who find that the appointing political party of federal judges influences judge decision making. It is worth emphasizing here, however, that we are measuring differences in judge responses to Sept. 11, 2001 along political affiliation and not differences in asylum approval along political affiliation.

1.7 Conclusion

Using the universe of asylum applications in the U.S. we investigated the effect of Islamic terrorism on granting rates for asylees from Muslim-majority countries in comparison to those from other countries. We find asylees from Muslim-majority countries were 3.2 percentage points less likely to be granted asylum in the aftermath of Sept. 11, 2001. For asylees from countries associated with the Sept. 11, 2001 attacks (Egypt, Lebanon, Saudi Arabia, United Arab Emirates) the effect is even more pronounced—they were 6–10 percentage points less likely to be granted asylum than those from other countries, however these estimates are relatively imprecise due to the small number of asylees from attack associated countries. We also find that the events of March 11, 2004 resulted in a 2.9 percentage point reduction in the likelihood that asylees from Muslim-majority countries were granted asylum. These results are robust to applicant and temperature controls, judge, city, month, and city-month fixed effects, as well as alternative estimation methods. Additionally, we do not find evidence of differences in the differential effects of Sept. 11, 2001 on asylees from Muslim-majority countries across judge political affiliation.

Our results are consistent with the emerging literature documenting the influence of emotions on decision making. These findings are of interest to policy makers in evaluating the performance of immigration judges and in preparing for potential consequences of terrorism for the justice system. Additionally, these findings are of interest in contributing to the research enumerating the consequences of terrorism and in particular the legacy of Sept. 11, 2001.

1.8 Figures

FIGURE 1.1: Average Monthly Relief Rates: Muslim-Majority Countries v. Other Applicants

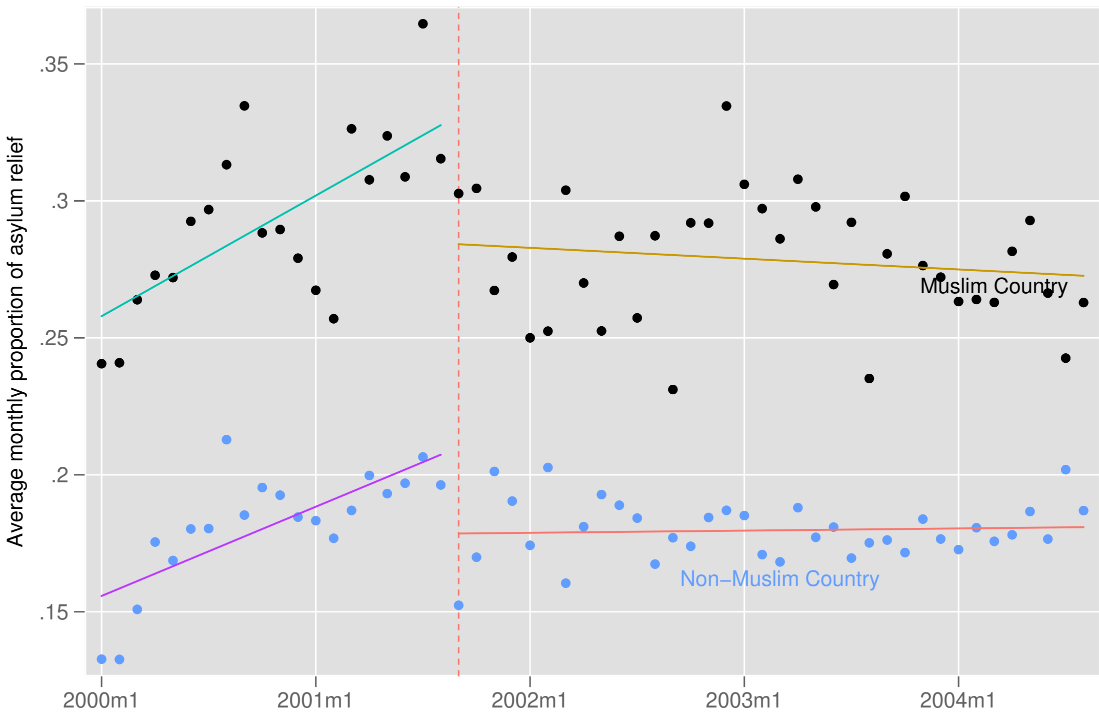
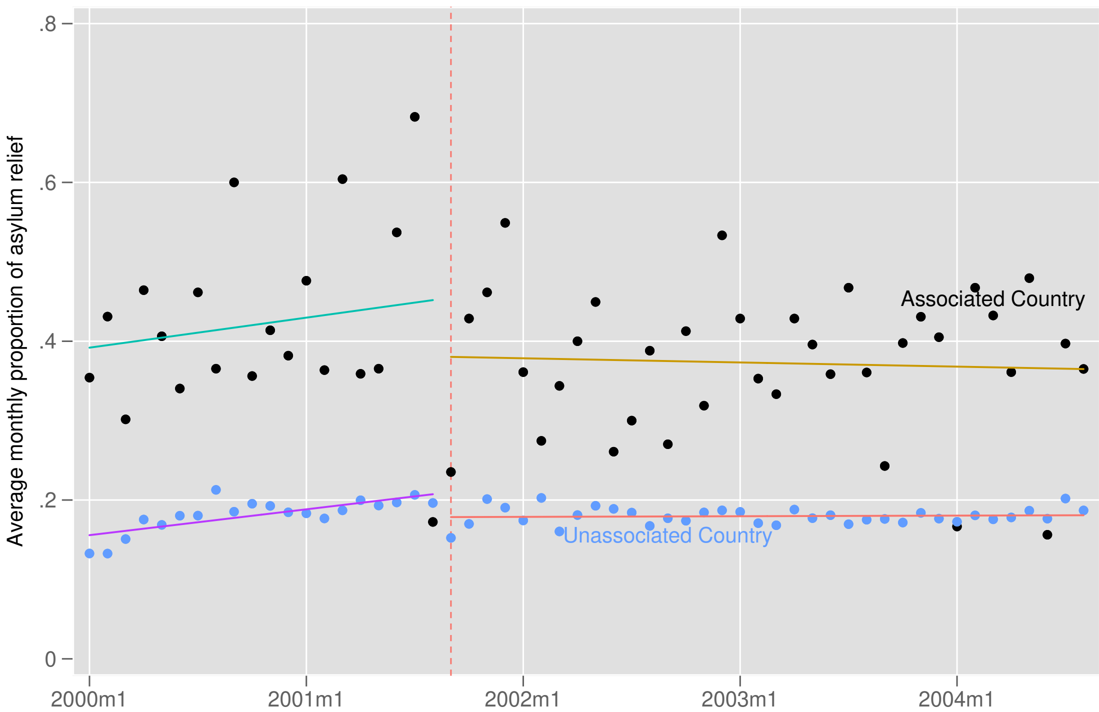


FIGURE 1.2: Average Monthly Relief Rates: Attack Associated Countries and Non-Attack Associated Countries



1.9 Tables

TABLE 1.1: Summary Statistics

	Mean	Std. Dev.
Outcomes		
Applicant Granted Any Relief	0.199	0.399
Applicants' Characteristics		
Asylee from Muslim-majority country	0.179	0.384
Asylee from Sept. 11, 2001 associated country	0.013	0.113
Defensive Application	0.292	0.455
Has Legal Representation	0.805	0.396
Location is DOC, Detention Center or Prison	0.082	0.274
Judges' Characteristics		
Judge Appointed During Democrat Presidency	0.708	0.455
Judge is Female	0.366	0.482
0–5 Year of Experience	0.436	0.496
6–10 Years of Experience	0.217	0.412
More 10 Years of Experience	0.347	0.476
Observations	269,270	

Note: Authors' calculations. See Section 1.4 for more details. The time period is January 2000–August2004.

TABLE 1.2: Effect of Sept. 11, 2001 on Any Relief Granted: Muslim-Majority Countries

	(1)	(2)	(3)	(4)	(5)
Asylee from Muslim-maj. country	0.112	0.131	0.121	0.121	0.120
	(0.048)	(0.048)	(0.034)	(0.034)	(0.034)
After Sept. 11, 2001	-0.020	-0.013	-0.008	-0.005	-0.009
	(0.008)	(0.009)	(0.008)	(0.011)	(0.011)
Muslim-maj. country ×		-0.038	-0.036	-0.036	-0.032
After Sept. 11, 2001		(0.014)	(0.012)	(0.012)	(0.011)
Observations	48,340	48,340	48,340	48,340	48,340
R-squared	0.012	0.012	0.174	0.174	0.186
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows OLS estimates of Equation (1). The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE 1.3: Effect of Sept. 11, 2001 on Any Relief Granted: Attack Associated Countries

	(1)	(2)	(3)	(4)	(5)
Asylee from associated country	0.249	0.303	0.252	0.251	0.248
	(0.056)	(0.048)	(0.034)	(0.034)	(0.034)
After Sept. 11, 2001	-0.014	-0.013	-0.009	-0.010	-0.017
	(0.009)	(0.009)	(0.008)	(0.012)	(0.012)
Associated country ×		-0.099	-0.064	-0.062	-0.060
After Sept. 11, 2001		(0.031)	(0.027)	(0.026)	(0.026)
Observations	39,882	39,882	39,882	39,882	39,882
R-squared	0.006	0.007	0.173	0.173	0.186
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows OLS estimates of Equation (1). The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE 1.4: Effect of March 11, 2004 on Any Relief Granted: Muslim-Majority Countries

	(1)	(2)	(3)	(4)	(5)
Asylee from Muslim-maj. country	0.092	0.102	0.047	0.048	0.048
	(0.060)	(0.061)	(0.035)	(0.036)	(0.036)
After March 11, 2004	0.006	0.010	0.008	0.003	0.004
	(0.007)	(0.008)	(0.006)	(0.010)	(0.009)
Muslim-maj. country ×		-0.022	-0.026	-0.027	-0.029
After March 11, 2004		(0.012)	(0.011)	(0.011)	(0.010)
Observations	64,393	64,393	64,393	64,393	64,393
R-squared	0.008	0.008	0.207	0.207	0.218
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows OLS estimates of Equation (1). The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is September 13, 2003 to August 31, 2004. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE 1.5: Effect of Sept. 11, 2001 on Any Relief Granted: Muslim-Majority Countries and Attack Associated Countries

	(1) 3 Months	(2) 6 Months	(3) 9 Months	(4) 12 Months	(5) 3 Months	(6) 6 Months	(7) 9 Months	(8) 12 Months
Asylee from Muslim-maj. country	0.123	0.120	0.107	0.101				
	(0.038)	(0.034)	(0.033)	(0.034)				
After Sept. 11, 2001	0.006	-0.009	-0.005	-0.006	0.010	-0.017	-0.003	-0.004
	(0.016)	(0.011)	(0.010)	(0.008)	(0.016)	(0.012)	(0.010)	(0.008)
Muslim country ×	-0.029	-0.032	-0.026	-0.022				
After Sept. 11, 2001	(0.013)	(0.011)	(0.015)	(0.018)				
Applicant from associated country					0.286	0.248	0.235	0.225
					(0.037)	(0.034)	(0.034)	(0.039)
Associated country ×					-0.099	-0.060	-0.045	-0.065
After Sept. 11, 2001					(0.048)	(0.026)	(0.012)	(0.010)
Observations	23,104	48,340	73,527	102,509	19,126	39,882	60,589	84,617
R-squared	0.191	0.186	0.179	0.172	0.195	0.186	0.182	0.177
Applicant Controls	✓	✓	✓	✓	✓	✓	✓	✓
Temperature	✓	✓	✓	✓	✓	✓	✓	✓
Judge FE	✓	✓	✓	✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓	✓	✓	✓
Month FE	✓	✓	✓	✓	✓	✓	✓	✓
City × Month FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table shows OLS estimates of Equation (1). The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. In columns 1 and 5, the time period is three months before and after Sept. 11, 2001. In columns 2 and 6, the time period is six months before and after Sept. 11, 2001. In columns 3 and 7, the time period is nine months before and after Sept. 11, 2001. In columns 4 and 8, the time period is twelve months before and after Sept. 11, 2001. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE 1.6: Effect of Sept. 11, 2001 on Any Relief Granted: Muslim-Majority Countries by Judges Appointed during Democrat v. Republican Presidencies

	(1) 3 Months	(2) 6 Months	(3) 9 Months	(4) 12 Months
Muslim-maj. country × After 9/11 × Democrat	0.016	0.002	0.019	0.003
	(0.038)	(0.025)	(0.025)	(0.026)
Observations	23,080	48,256	73,390	102,347
R-squared	0.118	0.112	0.104	0.101
Applicant Controls	✓	✓	✓	✓
Judge Controls	✓	✓	✓	✓
Temperature	✓	✓	✓	✓
City FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
City × Month FE	✓	✓	✓	✓

Note: This table shows OLS estimates of Equation (2). The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge controls include sex, experience, and political party of appointing administration. City fixed effects are included and temperature is the mean temperature on the day of the hearing from 6am–4pm. All pairwise interactions are included in the regressions, but omitted from the tables. Standard errors are clustered by country of origin and are shown in parentheses.

1.10 Appendices

TABLE A1: Test for Diverging Pre-Trends in Asylum Approval

	(1)
Muslim-maj. country × Time	0.0000551
	(0.0000786)
Observations	48,340
R-squared	0.1861
Applicant Controls	✓
Temperature	✓
Judge FE	✓
City FE	✓
Month FE	✓
City × Month FE	✓

Note: Coefficient is based on regressions of Equation (1) with the addition of a linear time trend and the interaction of the time trend and whether or not the asylee is from a Muslim-majority country. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A2: Test for Random Assignment of Cases to Judges

	(1)
Asylee from Muslim-maj. country	0.0000499
	(0.00000414)
Observations	269,270
R-squared	0.434
City FE	✓

Note: The dependent variable is a judge stringency measure created from leave-one-out mean grant rates for each judge. Coefficient presents results of regression of the judge stringency measure on a dummy variable that equals 1 if an asylee is from a Muslim-majority country and 0 otherwise. Court fixed-effects are included.

TABLE A3: Test for Change in Likelihood of Muslim Cases of Judges

	(1)
After Sept. 11, 2001	0.012
	(0.013)
Observations	48,340
R-squared	0.130
Applicant Controls	✓
Temperature	✓
Judge FE	✓
City FE	✓
Month FE	✓
City × Month FE	✓

Note: The dependent variable is a binary variable that equals 1 if an asylee is from a Muslim-majority country. Coefficient presents results of regression of this Muslim-country dummy on a dummy variable that equals 1 if the case is heard after Sept. 11, 2001 and 0 otherwise. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm.

TABLE A4: Full Estimation of the Effect of Sept. 11, 2001 on Any Relief Granted: Six Month Window

	(1)
Asylee from Muslim-maj. country	0.120
	(0.034)
After Sept. 11, 2001	-0.009
	(0.011)
Muslim-maj. country × After Sept. 11, 2001	-0.032
	(0.011)
Defensive Application	-0.005
	(0.022)
Has Legal Representation	0.191
	(0.035)
Location is DOC, Detention centre, Jail, or Prison	-0.005
	(0.024)
Temperature	0.008
	(0.006)
Observations	48,340
R-squared	0.186
Applicant Controls	✓
Temperature	✓
Judge FE	✓
City FE	✓
Month FE	✓
City × Month FE	✓

Note: This table shows OLS estimates of Equation (1). The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A5: Logit Estimation of the Effect of Sept. 11, 2001 on Any Relief Granted: Six Month Window

	(1)	(2)	(3)	(4)	(5)
Asylee from Muslim-maj. country	0.101	0.113	0.108	0.108	0.108
	(0.037)	(0.035)	(0.027)	(0.027)	(0.027)
After Sept. 11, 2001	-0.020	-0.014	-0.010	-0.006	-0.012
	(0.008)	(0.010)	(0.008)	(0.010)	(0.012)
Muslim-maj. country ×		-0.026	-0.027	-0.026	-0.022
After Sept. 11, 2001		(0.014)	(0.011)	(0.010)	(0.009)
Observations	48,350	48,350	48,070	48,070	47,139
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows Logit estimates of Equation (1). Marginal effects are reported. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A6: Probit Estimation of the Effect of Sept. 11, 2001 on Any Relief Granted: Six Month Window

	(1)	(2)	(3)	(4)	(5)
Asylee from Muslim-maj. country	0.102	0.116	0.108	0.108	0.108
	(0.039)	(0.037)	(0.027)	(0.027)	(0.027)
After Sept. 11, 2001	-0.020	-0.014	-0.010	-0.006	-0.011
	(0.008)	(0.009)	(0.008)	(0.010)	(0.012)
Muslim-maj. country ×		-0.028	-0.028	-0.027	-0.022
After Sept. 11, 2001		(0.014)	(0.011)	(0.011)	(0.009)
Observations	48,350	48,350	48,070	48,070	47,139
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows Probit estimates of Equation (1). Marginal effects are reported. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A7: Placebo Test using Sept. 11, **2000**: Six Month Window

	(1)	(2)	(3)	(4)	(5)	(6)
Asylee from Muslim-maj. country	0.107	0.095	0.095			
	(0.048)	(0.035)	(0.035)			
After Sept. 11, 2000	0.008	0.019	0.038	0.008	0.017	0.043
	(0.006)	(0.006)	(0.015)	(0.006)	(0.006)	(0.015)
Muslim-maj. country ×	-0.009	-0.015	-0.017			
After Sept. 11, 2000	(0.014)	(0.012)	(0.012)			
Applicant from associated country				0.219	0.190	0.192
				(0.108)	(0.067)	(0.068)
Associated country ×				0.022	0.025	0.022
After Sept. 11, 2000				(0.021)	(0.012)	(0.013)
Observations	47,672	47,672	47,672	39,342	39,342	39,342
R-squared	0.010	0.161	0.162	0.006	0.167	0.167
Applicant Controls		✓	✓		✓	✓
Temperature		✓	✓		✓	✓
Judge FE		✓	✓		✓	✓
City FE		✓	✓		✓	✓
Month FE			✓			✓

Note: This table shows OLS estimates of Equation (1) for a placebo analysis. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2002 to March 10, 2003. Applicant controls include application type, detention status, and legal representation. Judge, city, and month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A8: Placebo Test using Sept. 11, **2002**: Six Month Window

	(1)	(2)	(3)	(4)	(5)	(6)
Asylee from Muslim-maj. country	0.090	0.068	0.069			
	(0.049)	(0.032)	(0.033)			
After Sept. 11, 2002	0.000	-0.008	-0.002	0.000	-0.006	0.006
	(0.006)	(0.005)	(0.010)	(0.006)	(0.005)	(0.011)
Muslim-maj. country ×	0.023	0.006	0.003			
After Sept. 11, 2002	(0.018)	(0.017)	(0.017)			
Asylee from associated country				0.175	0.141	0.141
				(0.073)	(0.046)	(0.047)
Associated country ×				0.024	-0.007	-0.009
After Sept. 11, 2002				(0.030)	(0.027)	(0.026)
Observations	58,914	58,914	58,914	49,132	49,132	49,132
R-squared	0.010	0.197	0.197	0.004	0.206	0.207
Applicant Controls		✓	✓		✓	✓
Temperature		✓	✓		✓	✓
Judge FE		✓	✓		✓	✓
City FE		✓	✓		✓	✓
Month FE			✓			✓

Note: This table shows OLS estimates of Equation (1) for a placebo analysis. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 200 to March 9, 2001. Applicant controls include application type, detention status, and legal representation. Judge, city, and month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A9: Placebo Test using Sept. 11, **2003**: Six Month Window

	(1)	(2)	(3)	(4)	(5)	(6)
Asylee from Muslim-maj. country	0.104	0.048	0.048			
	(0.055)	(0.031)	(0.030)			
After Sept. 11, 2003	-0.001	-0.007	0.002	-0.001	-0.009	-0.002
	(0.006)	(0.007)	(0.011)	(0.006)	(0.008)	(0.011)
Muslim-maj. country ×	-0.000	0.003	0.002			
After Sept. 11, 2003	(0.017)	(0.016)	(0.016)			
Asylee from associated country				0.227	0.174	0.174
				(0.061)	(0.030)	(0.029)
Associated country ×				-0.019	-0.026	-0.026
After Sept. 11, 2003				(0.074)	(0.054)	(0.054)
Observations	69,792	69,792	69,792	58,923	58,923	58,923
R-squared	0.010	0.187	0.187	0.004	0.191	0.191
Applicant Controls		✓	✓		✓	✓
Temperature		✓	✓		✓	✓
Judge FE		✓	✓		✓	✓
City FE		✓	✓		✓	✓
Month FE			✓			✓

Note: This table shows OLS estimates of Equation (1) for a placebo analysis. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2003 to March 9, 2004. Applicant controls include application type, detention status, and legal representation. Judge, city, and month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A10: Madrid Placebo Test using March 11, **2003**: Six Month Window

	(1)	(2)	(3)	(4)	(5)
Asylee from Muslim-maj. country	0.109	0.114	0.065	0.066	0.068
	(0.052)	(0.050)	(0.028)	(0.028)	(0.028)
After March 11, 2003	-0.003	-0.001	0.003	0.009	0.005
	(0.006)	(0.007)	(0.006)	(0.011)	(0.010)
Muslim country ×		-0.011	-0.010	-0.012	-0.016
After March 11, 2003		(0.018)	(0.015)	(0.015)	(0.013)
Observations	66,089	66,089	66,089	66,089	66,089
R-squared	0.010	0.010	0.186	0.186	0.196
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows OLS estimates of Equation (1) for a placebo analysis of the March 11, 2003 attacks in Madrid. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is Sept. 12, 2002 to Sept. 7, 2003. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A11: Estimation of the Effect of Sept. 11, 2001 on Any Relief Granted excluding Withdrawals and Dismissals: Six Month Window

	(1)	(2)	(3)	(4)	(5)
Asylee from Muslim-maj. country	0.122	0.143	0.120	0.120	0.120
	(0.048)	(0.052)	(0.032)	(0.032)	(0.032)
After Sept. 11, 2001	-0.024	-0.014	-0.005	0.022	-0.005
	(0.012)	(0.011)	(0.009)	(0.016)	(0.009)
Muslim-maj. country ×		-0.043	-0.044	-0.044	-0.044
After Sept. 11, 2001		(0.023)	(0.017)	(0.017)	(0.017)
Observations	25,819	25,819	25,819	25,819	25,819
R-squared	0.011	0.011	0.236	0.236	0.236
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows OLS estimates of Equation (1) excluding withdrawn and dismissed cases. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

TABLE A12: Varying the Threshold for Muslim-Majority Country Full Estimation of the Effect of Sept. 11, 2001 on Any Relief Granted: Six Month Window

	(1)	(2)	(3)	(4)	(5)
Asylee from 75% Muslim-maj. country	0.093	0.110	0.111	0.111	0.108
	(0.048)	(0.048)	(0.037)	(0.037)	(0.037)
After Sept. 11, 2001	-0.020	-0.015	-0.009	-0.006	-0.010
	(0.008)	(0.009)	(0.008)	(0.011)	(0.010)
75% Muslim-maj. country ×		-0.035	-0.035	-0.035	-0.028
After Sept. 11, 2001		(0.016)	(0.013)	(0.013)	(0.013)
Observations	48,340	48,340	48,340	48,340	48,340
R-squared	0.007	0.007	0.171	0.171	0.183
Applicant Controls			✓	✓	✓
Temperature			✓	✓	✓
Judge FE			✓	✓	✓
City FE			✓	✓	✓
Month FE				✓	✓
City × Month FE					✓

Note: This table shows OLS estimates of Equation (1). The Muslim-majority country definition is changed from 50% to 75%. The unit of observation is a case. The dependent variable is a dummy that equals one if asylum is granted and zero otherwise. The time period is March 15, 2001 to March 9, 2002. Applicant controls include application type, detention status, and legal representation. Judge, city, month, and city-month fixed effects are included. Temperature is the mean temperature on the day of the hearing from 6am–4pm. Standard errors are clustered by country of origin and are shown in parentheses.

Chapter 2

Policing for whom? Officer-involved shootings and police legitimacy in Chicago

2.0.1 Abstract

This paper explores the effects of officer-involved shootings (OIS) on perceptions of police legitimacy and how those perceptions shape civilian action and behavior. I focus on Chicago, leveraging a rich dataset containing the universe of officer firearm discharge, and examine the effects of OIS on civilian reported crime. To identify the effects of OIS on civilian reported crime, I exploit the randomness of subjects' injury status from firearm discharges after adjusting for a broad set of officer, subject, and incident characteristics. I provide evidence that after accounting for these factors, whether or not a subject is injured is as good as random because police are trained to shoot to incapacitate the subject rather than to wound, disarm, or deter. Comparing districts with OIS where subjects are injured and districts with OIS where subjects are not injured, I find that reported crime falls by 2.8% following OIS resulting in injury. I find no differential effects for Black subjects, Black districts, White-Black officer-subject pairs, or fatally injured subjects. Instead, I find even larger reductions when subjects are unarmed. I argue that these results are consistent with a model in which concerns about procedural justice are the primary determinants of citizens' perceptions of police legitimacy.

2.0.2 Thanks

Special thanks go to Louis-Philippe Béland, Abel Brodeur, Aaron Chalfin, Jamein Cunningham, Jennifer Doleac, Jason Garred, Rob Gillezeau, Myra Mohnen, Louis-Philippe Morin, and participants at CEA for comments and suggestions. Additional thanks to the Invisible Institute for providing data. Any errors are my own.

2.1 Introduction

From the shootings of Michael Brown and Philando Castile to the recent deaths of George Floyd and Duante Wright, it is perhaps no surprise that U.S. trust in police is at an all-time low in Gallup's 27 years tracking the question (Ortiz, 2020). The images, videos, and descriptions of police shootings and use of force have sparked numerous protests, spawned the national movement Black Lives Matter, and elevated calls to defund the police.[1]

The public police violence tracking project Mapping Police Violence estimates that as of August 1, 2021 police in the U.S. have killed 564 people—over 2.5 deaths per day—while in 2020 there were 1,126 deaths (∼3 deaths per day). These data indicate that Black Americans are disproportionately killed by police relative to their share of the population (28% of deaths compared to 13% of the population) (Mapping Police Violence, 2021). Pew Research found that Blacks are about half as likely as Whites to view police use of force or treatment of racial groups positively and were 30 percentage points less likely than Whites to believe that police do a good job protecting people from crime (DeSilver, Lipka, and Fahmy, 2020).

On top of the staggering human costs of police violence, there are two additional costs that bear consideration, especially from an economic perspective: the equity and efficiency of public safety provision by the state. Police violence may cause inequities in the provision of public safety if civilians of color, and specifically Black civilians, are less likely to report crime to police or cooperate with investigations because of fears that police are not competent or that their involvement might make matters worse. Such changes in behavior may mean that police violence itself generates subsequent crime and instability in public safety (including differences across racially concentrated areas), elevating the priority of measures to reduce police use of force.[2] Lower clearance rates (the share of cases deemed closed) following police violence imply a less efficient police force and create incentives for policy makers to reallocate resources away from police and towards other public safety policies and services.

In this paper, I examine how officer-involved shootings (OIS) influence perceptions of police legitimacy and how those perceptions shape civilian action and behavior. I first show suggestive evidence that negative sentiment in tweets about police spikes after OIS using two case studies–the shootings of Justine Damond (a 40 year old unarmed white woman) and Stephon Clark (a 22 year old unarmed Black man). Using all US tweets containing "cop", "cops", or "police" I classify the words in tweets as positive or negative. In the two weeks prior to these incidents, about 7% of words used in tweets about police are negative but this share spikes to over 15% after news stories of Damond's shooting begin appearing and to 9% after Clark's shooting. While not causal, these case studies provide suggestive evidence that perceptions of police among some Americans are responsive to officer-involved shootings.

In order to say something about whether or not these perceptions influence actions, I then focus on Chicago, leveraging an incredibly rich dataset containing the universe of officer firearm discharges, and examine the effects of OIS on civilian reported crime. To identify the effects of OIS on civilian reported crime, I exploit the randomness of subjects' injury status from firearm discharges after adjusting for officer, subject, and incident characteristics (e.g. officer experience and weapon discharge history, race, age, gender; lighting and weather conditions; and subject weapon status, race, age, gender). I argue that after accounting for these factors, whether or not a subject is injured is as good as random because police are trained to shoot to incapacitate the subject rather than to wound, disarm, or deter. I then compare the rate of reported crime in police districts where there was an OIS with

[1] According to data from Count Love, a public protest tracking initiative, 2020 saw over 8,000 protests for racial justice in the U.S. (Count Love, 2021).

[2] In 2021, the largest 50 cities in America directed 13.7% of their general expenditures towards law enforcement (Akinnibi, Holder, and Cannon, 2021).

an injury to police districts where there was an OIS without an injury before and after the OIS. This identification strategy is appealing since OIS with and without subject injury have similar officer, subject, and incident characteristics and these characteristics struggle to predict subject injury.

Comparing districts with OIS where subjects are injured and districts with OIS where subjects are not injured, I find that reported crime falls by about 2.8% following OIS resulting in injury. These findings are consistent with the idea that OIS with and without subject injuries differ in two key dimensions. First, OIS with subject injuries may receive more media coverage than OIS without injuries. Second, OIS with subject injuries have the possibility of resulting in death whereas OIS without subject injury do not. In other words, OIS with subject injuries are more salient than OIS without subject injuries, which may explain the lack of change in crime reporting for districts with OIS where subjects are uninjured. When including TASER discharges I do not find any reduction in reported crime, further emphasizing the importance of salience.

Turning to which types of OIS might be driving this result, I find no differential effects for Black subjects, Black districts, or White-Black officer-subject pairs. I also do not find differential effects when subjects are fatally injured, though these estimates are much less precise than the others. Instead, I find large additional reductions in reported crime when subjects are unarmed. These results together suggest that it is the process rather than outcomes (i.e. procedural justice rather than distributive justice) that matters more for determining perceptions of police legitimacy and guiding citizen behavior. My findings support the idea that, irrespective of CPD policy, the shooting of an unarmed citizen is seen as procedurally unjust and damages not only perceptions of police legitimacy but also public engagement with law enforcement.

My work makes several contributions to understanding OIS and their consequences. First, I extend our knowledge about the determinants of police legitimacy beyond survey settings and examine manifestations of citizen behavior. Despite police legitimacy owning an extensive literature in criminology and sociology, it is mainly drawn from surveys and there has been relatively little well-identified empirical work done. A key challenge is the measurement of police legitimacy. Existing studies largely rely on surveys (see, for example, Kochel, 2015a, Kochel, 2015b, or White, Weisburd, and Wire, 2018) but these have two significant limitations: they are relatively infrequent, and responses do not map cleanly to civilian behaviors (e.g. Bobo and Thompson, 2006 find that despite differences in attitudes towards police by Black and White respondents, their proposed actions were the same). My approach using Twitter data gives insight about perceptions while the reported crime data maps directly to civilian actions.

My work also expands on existing studies by moving beyond an event study framework, examining many OIS incidents, exploring which aspects of OIS might generate differential effects, and which types of crimes might be affected. Recent work examining the effects of OIS on crime reporting (Desmond, Papachristos, and Kirk, 2016, Ang et al., 2021) and sentiment about police Oglesby-Neal, Tiry, and Kim, 2019 find lower crime reporting and more negative sentiments but use only a small number of high-profile cases that garnered media attention (furthermore there is debate about whether or not the crime reporting effect is genuine, see Zoorob, 2020 and Desmond, Papachristos, and Kirk, 2020).

Last, a common approach in studies examining the impacts of OIS and the "Ferguson effect" is to adjust for the rate of crime, as increases in crime could stretch police resources more thinly and explain lower clearance rates or increases in homicides. However, as the rate of crime is determined both by police behaviors (which this strand of research suggests are influenced by OIS) and police legitimacy through civilian reporting behavior, then the crime rate is actually an outcome and its inclusion introduces bias when estimating the effects of OIS on clearance rates and crime reporting. I leverage the acoustic gunshot detection system ShotSpotter to measure gun violence directly to examine how adjusting for underlying crime influences the effects of OIS on crime reporting.

The remainder of this paper is structured as follows. In Section 2.2, I detail how OIS can influence police legitimacy and reported crime. Section 2.3 describes the dataset and provides descriptive statistics. Section 2.4 presents the empirical strategy and required identifying assumption. Section 2.5 presents and discusses results. Section 2.6 concludes.

2.2 Framework

The theoretical effects of OIS with injuries on civilian crime reporting are unclear. Several channels are at work.

2.2.1 Police Legitimacy and Public Support of Police

One channel that OIS can influence reporting behaviour through is perceptions of police legitimacy. In order to operate efficiently, legal institutions need to draw on feelings of obligation and responsibility from members of the public to facilitate cooperation and compliance. Legitimacy is the belief by people that some authority deserves to be obeyed. When the public views institutions as legitimate, they will voluntarily comply with directives. Police legitimacy, therefore, is the belief by the public that they should defer to the police and assist with crime prevention.

Broadly speaking there are two main approaches to thinking about the determinants of police legitimacy. The outcomes based approach posits that police gain acceptance when they can credibly sanction rule breakers; effectively control crime; and fairly distribute services across the public. The procedural justice approach links legitimacy to the public's evaluations about the fairness of process that the police use to make their decisions and exercise their authority (Sunshine and Tyler, 2003, Tyler and Huo, 2002). There is an extensive literature in criminology and sociology composed mainly of surveys examining the relative importance of these models describing how the public evaluates police legitimacy. This qualitative work finds that youth in high-crime areas have negative views of police often driven by repeated harassment and police misconduct and respondent narratives frequently centre around the fairness and justice of interactions rather than outcomes (Berg et al., 2016, Gau and Brunson, 2010, Wolfe et al., 2016). Personal interaction is not the only way to generate these perceptions of police legitimacy, the experience of others, often family members, matters as well and exhibits a similar focus on fairness and justice (Brunson, 2007, Carr, Napolitano, and Keating, 2007).

Studies have also found that local conditions matter, with structurally disadvantaged neighborhoods being more likely to have lower perceptions of police legitimacy (Kirk and Papachristos, 2011). For example, McCarthy, Hagan, and Herda, 2020 find that complaints filed against CPD members disproportionately originate in racially segregated neighborhoods and that previous measures of legitimacy predict current complaint behaviour. Additionally, research has found that repeated exposure to media reports of police abuse is strongly positively related to perceptions of police misconduct (Kaminski and Jefferis, 1998, Weitzer, 2002, Weitzer and Tuch, 2004). That is, the media is mediating perceptions of police legitimacy.

How does police legitimacy interact with OIS and reporting behavior? Even when an OIS is deemed acceptable within the police department, the public may believe that the incident should have been addressed without the use of lethal force. This is an erosion of procedural justice which we expect to lower perceptions of police legitimacy. Kwak, Dierenfeldt, and McNeeley, 2019 use victimization survey to explore the relationship between procedural justice and crime reporting, finding that lower perceptions of procedural justice was associated with a lower probability of reporting crime by victims. In Anderson, 1999's seminal ethnographic depiction of Philadelphia, he writes "Residents sometimes fail to call the police because they believe that the police are unlikely to come or, if they do come, may even harass the very people who called them." (p.321). Poor views of police legitimacy may mean that citizens do not call the police, worrying that officers may escalate a situation or engage in misconduct.

Circling back to the discussion about procedural justice, we might expect larger declines in reporting for OIS with unarmed subject as the disproportionate use of force should have a large negative impact on perceptions of police legitimacy. However, the distributive justice approach is often seen in popular media framing of OIS through a racial lens and means that we might expect to see larger effects when the victim is Black, when the shooting occurs in a majority Black district, or when the officer is White and subject is Black.

The elasticity of crime reporting to OIS might be different depending on the severity of the crime or the level of discretion that civilians have in reporting that crime. We might expect that more serious crimes like homicide or burglary require an immediate response and therefore have an inelastic response to OIS–those crimes were always going to be reported. However, civilians may believe that less serious crimes, like narcotics or liquor law violations, do not require alerting the police because of the lack of urgency or immediate threats to safety. While this is true in general, may be especially true in the aftermath of an OIS where citizens worry that alerting police could escalate a situation and result in another OIS. In this scenario, the expected elasticity of reporting to OIS should be high. Lastly, in the absence of legitimate police, violence becomes a problem solving tool used to mete out justice. As such, falling police legitimacy from an OIS could result in upward pressure on crime and in particular, violent crime.

2.2.2 Economic Model of Crime

Another channel is related to changes in the underlying crime rate. Becker, 1968 introduced an economic theory of crime where individuals make decisions about committing a crime based on comparing the expected costs and benefits of doing so. Those costs are determined by and increasing in the probability of apprehension and the severity of the punishment. In this context, and OIS may increase the perceived cost of engaging in crime by making this cost more salient. Even if, traditionally, an optimizing agent would already take this cost into account, behaviour economics and cognitive psychology have identified biases and heuristics that influence decision making. In particular the availability heuristic might be at work here, where the ease to which an individual can recall an event influences the perceived probability that event occurs. If we consider that the severity of punishment includes death during apprehension, then an OIS should increase the cost of engaging in crime and create a disincentive to engage in criminal activity.

On the other hand, the notion of the Ferguson effect is based on the premise that police scale back proactive activities in the aftermath of an OIS because they are concerned about the perceptions of their actions under this additional scrutiny (Gaston, Cunningham, and Gillezeau, 2019). The Ferguson effect is a proposed explanation for two phenomenon: 1) following OIS there is seen to be an increase in homicides and 2) following OIS there is a reduction in case clearance rates. In the context of the Becker model, a reduction in this police activity should lower the probability of apprehension and thus create an incentive to engage in criminal activity. And specifically, the Ferguson effect indicates that we should expect an increase for violent crime. However, there is an emerging consensus that this depolicing does not result in increased homicides.

Some have argued that the Becker model is a good description of behaviour for property and non-violent crimes but that it falls short in explaining violent crime behaviour because of the rationality assumption (see for example,). Under the assumption that violent crime is motivated more by proximity, opportunity, and emotions we might expect that a reduction in the probability of conviction would not alter the rate of violent crime and that instead we should see an increase instead in only non-violent crime.

Taken together this means that the direction of crime following an OIS is ambiguous and the direction of crime reporting following an OIS is likely negative. Even if there is a decline in crime reporting following an OIS, it could simply reflect a lower level of crime rather than a decline in perceived police legitimacy and crime reporting behaviour. In Section 2.5, I attempt to disentangle these two effects by using data about gun violence in Chicago measure by acoustic gunshot sensors (ShotSpotter).

2.2.3 Twitter and sentiment about police

Can perceptions of police change after OIS? To answer this question I turn to Twitter data and analyze the sentiment and emotional content of police-related tweets and how that evolves before and after two OIS case studies. As noted in 2.1, there is evidence that public perceptions of police change in response to OIS—both using survey data (Kochel, 2015a) and using Twitter data (Oglesby-Neal, Tiry, and Kim, 2019). It is important to note that overall Twitter is not nationally representative. Survey data from PEW Research indicate that approximately 22% of the U.S. population use Twitter and these users are on average younger, more likely to vote Democrat, earn higher incomes and be more likely to agree with statements indicating the existence of social inequalities rooted in race ($\sim$ 10 percentage points) and gender ($\sim$ 6 percentage points) (**wojcikSizingTwitterUsers2019**). These users however, are just as likely as the general U.S. population to be White, Black, or Hispanic or to be men or women. And in particular, Twitter users are 5 percentage points less likely than the general population to say they are "very attached" to their local community, which may suggest that changes in Twitter content may be less influenced by local events. These case studies are meant to illustrate the possibility that the sentiment of tweets about police can change in response to OIS rather than to make a causal claim about perceptions of police changing on average in America in response to OIS.

I use Twitter's Academic Research product track which allows full access to Twitter's archives through their API. I then searched for tweets that were tagged as being from the United States during 2017–2019 mentioning "cops", "cop", or "police" to build a corpus of tweets. I did not search for tweets specifically related to incidents of police brutality or officer-involved shootings/fatalities as I am interested in the sentiment of discourse about police overall in response to these incidents rather than discourse specifically about these incidents. I then remove stop words (commonly used function words such as "the" and "is") and separate tweets into individual words before applying the NRC Word-Emotion Association Lexicon (aka EmoLex) introduced by Mohammad and Turney, 2013 which allows for binary sorting of words as positive and negative as well as their association with eight different emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust).[3]

In particular, I present two case studies of high profile police shootings in 2017 and 2018. Justine Damond was an unarmed 40 year old, White, Australian-American living in Minneapolis, Minnesota. Shortly after 11:30pm on July 15, 2018 she called 911 to report a possible assault taking place behind her house. Officers Mohamed Noor (33 year old, Black, Somali-American with 21 months of experience) and Matthew Harrity (25 year old, White, American with 1 year of experience) responded to the scene and determined the alley was empty and the scene was safe. According to testimony, the officers then heard a loud noise and Damond appeared immediately outside the police vehicle. Harrity drew is weapon but did not fire and Noor fired a shot through the vehicle's open window that fatally injured Damond (Park, Grinberg, and Yan, 2017).

Stephon Clark was an unarmed 22 year old Black, American living in Sacremento, California. Shortly before 9:30pm pn March 18, 2018 he was shot by Sacremento police officers Terrence Mercadal (Black, American with 3 years of experience) and Jared Robinet (White, American with 4 years experience) who were responding to a call than someone was breaking nearby car windows. After being directed to the location of a suspect seen breaking a window with a tool bar by deputies in a helicopter, officers confronted a man who was later identified as Clark. Officers told him to show his hands and he fled to the back of the property before turning towards the officers with an object in his hands. Officers then fired on Clark, fatally injuring him.

Figures 1 and 2 show the average share of words per police-related tweet registering as positive, negative, and as each emotion for the entire month each of the shootings of Damond and Clark occurred in, respectively. In both figures overall sentiment for the month is more negative than positive and the most common emotions are fear and trust.

Figure 3 plots the average share of words in each police-related tweet that are positive or negative for a given day during the month of July 2017 and 4 does the same for March 2018. The horizontal axis plot the days since the Damond and Clark shootings, respectively. We can see that in the weeks before each shooting negative sentiment in police tweets was about 7% of words and that after each shooting negative sentiment in tweets about police spikes. For the Clark shooting negative sentiment is up to around 9% and for the Damond shooting it is up to nearly 15%.[4]

These results indicate that among Twitter users, the language used in tweets about police is responsive to OIS. This does not necessarily mean that the American population is changing the way they view police but it is consistent with the existing literature suggesting that OIS shape the way citiens view police.

[3]The patterns are consistent when using other lexicons that assign words values from -5 to 5 instead of a binary classification.

[4]The large uptick right before the Clark shooting could be the result of a Washington Post article about the continued deaths of Black Americans at the hands of police

The remainder of this paper examines the question, "if citizens perceptions about police change, does that translate into changes in behavior?" This translation of how OIS influence perceptions about police into how OIS influences cooperation with police could result from the changes in Twitter being representative of the general public and capturing a broad based change in cooperation with police, or with the notion that Twitter users are responsible for more crime reporting than non-Twitter users (assuming that the changes in perceptions are soleley among Twitter users). The existing literature that uses survey data (again see Kochel, 2015b, Kochel, 2015a) also finds changes in perceptions towards police following OIS is suggestive of the former. Also suggestive of the former, is the idea that Twitter users are expressing these changes in perception in response to traditional media coverage of these OIS—in which case, these OIS are also more salient to non-Twitter users who consume the accompanying media coverage and subsequently update their attitudes and perceptions of police.

2.3 Data

2.3.1 Reported Crime Incidents

Data on reported crime incidents covers the 2004–2019 period and comes from the city of Chicago's open data portal. These reports come from the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting system and are censored at the block level. It is important to note that the data contained are preliminary and based on the incidents reported by third parties to the police department and may be unverified. That is, they represent the information as presented by those calling in requests for police service. The dataset contains the date, time, and location (censored at the block level) of the incident; the police beat and district the incident occurred in; the type of offense; and indicators for arrest and domestic violence.

The Chicago Police Department, reports its data to the FBI using the Uniform Crime Reporting Program. This system follows a hierarchy rule where if multiple offenses occur, only the most serious offense is recorded. The hierarchy is as follows: homicide, rape, robbery, aggravated assault, burglary, theft, motor vehicle theft, and lastly arson. The hierarchy rule already been applied to the reported crime data so that they conform with the UCR handling. This means that the data are an undercount though the FBI's examination of the effect of multiple reporting incidents indicates that they are a small share of overall incidents ($\sim$10%) and when UCR data is compared to its incident based successor, National Incident Based Reporting System (NIBRS), the undercounting appears to be relatively small ($\sim$2% increase in overall crimes if allowing up to 10 offenses per incident). I then use the Illinois Uniform Crime Reporting (IUCR) codes, to separate these crimes into Part I UCR offenses (which I refer to as "serious" crimes), Part I UCR violent offenses (which I refer to as "violent" crimes), and Part II UCR offenses (which I refer to as "less serious" crimes). Serious crimes are homicide, rape, robbery, aggravated assault, burglary, theft, motor vehicle theft, and arson. Violent crimes are homicide, rape, robbery, and aggravated assault. Less serious crimes include offenses such as vandalism, simple assault, weapons violations, disorderly conduct, drunkenness, narcotics and liquor laws, and disturbing the peace.

Table 1 presents descriptive statistics over my sample, with the first grouping of variables corresponding to crime reporting. On average, there are about 47 reported crimes per district-day. The majority of these are less serious crimes with about 28 reported each district-day (about 60% of all reported crimes). There are approximately

18 serious crimes reported per district-day and the majority of those are non-violent (about 84% of serious crimes).

Figure 6 plots the daily average reported crimes per district over the sample period. Two key features present themselves from this figure. First, reported crime has been on a downward trend over the sample period. Second, reported crime exhibits patterns of seasonality with local peaks during the summer. I address these in my empirical approach by including month, year, and district-year fixed effects. Figure 7 plots the distribution of reported crimes. Reported crimes are not particularly skewed, nor do they have an excess number of zeros. Still, in my analysis I transform the measures of reported crime using the natural log + 0.1 for ease of interpretation as semi-elasticities, though tables report the mean of the dependent variable in levels to help provide context. My main results are robust to estimating without the log transformation.

2.3.2 Officer Involved Shootings

Data detailing OIS covers 2004–2019 and comes from the Invisible Institute Citizens Police Data Project. This project used Freedom of Access to Information requests to obtain CPD records including tactical response reports. Tactical response reports (TRRs) are forms that Chicago Police officer must file when an officer discharges a firearm, impact munition, TASER, pepper spray or another chemical weapon or after qualifying use of force incidents.[5]

The TRRs include the location, date and time of the incident; officer information such as age, race, gender, appointment date, rank, duty status at the time of the incident, the type of firearm/weapon discharged; subject information such as age, race, gender, the charges against; weapon type (if any); and incident information such as the lighting and weather conditions, whether the OIS occurred inside or outside, whether or not subject is actively or passively resisting, used deadly force against an officer, and which party fired first.[6] I restrict my sample to include only OIS involving TASERs and firearms but use the entire pool of TRR incidents to create a variable that counts the total number of TRR incidents for a given officer. I explore OIS where TASERs are discharged because they are still weapons that can result in injury (∼42% of OIS involving TASERs result in injury in my sample) or death. If citizens being injured by police in OIS influences perceptions of police legitimacy and subsequent crime reporting behavior, it may not only be firearms that matter but TASERs as well.

Table 1 provides descriptive statistics. On average, these CPD members are 37 years old, White (52%), male (89%), officers (74%) in uniform (76%) with 10 years of experience and nine prior TRR incidents. These officers discharge TASERs 85% of the time and handguns in the remaining incidents (there are a handful of incidents where officers discharge a long gun but these are exceedingly uncommon). According to the Office of the Ispector General for the city of Chicago, in 2017 51% of officers were White, 23% Hispanic, and 21% were Black while 77% were male and 23% were female. The average CPD officer was 42 years old with 14.7 years of experience. So,

[5]A TRR must be filed for any use of force incidents when: a subject is injured or alleges injury; a subject actively resists; a subject whose actions are "aggressively offensive" (with or without weapons) or who is threatening immediate use of of force that will likely result in injury; physically obstructing an officer; or an assault, threat of physical attacks, or physical attack against an officer even if the officer is not injured.

[6]An example TRR is provided in the Appendix.

CPD officers who discharge their firearms or TASERs tend to be younger, more inexperienced, more male, less Black, and more Hispanic than the average CPD officer. Subjects are, on average, 30 years old, Black (77%), males (93%) who are unarmed (76%), and injured (55%). When armed, subjects most commonly have a hand gun (11% of all incidents). Over the course of an OIS, subjects are passively (3.41 actions) and actively (2.75 actions) resisting officer directives.

Subjects use deadly force in 17% of incidents, have assault charges in 63%, domestic charges in 14%, drug charges in 19% and weapons charges in 13%. OIS, on average, occur outside (75%) with clear weather (88%) and either under artificial light (55%) or daylight (31%). These data provide rich controls not only for our officer and subject characteristics but also allow me to deal with concerns about visibility and the danger of the incident which may be related to the injury status of the victim. Further discussion of the relationship between injury status and these covariates can be found in Section 2.4.

2.3.3 ShotSpotter

In order to test whether or not accounting for the underlying crime rate matters, I turn to ShotSpotter data obtained from the city of Chicago. ShotSpotter is an acoustic gunshot detection technology that makes use of sensors and microphone equipment to detect presence and locations gunshots. Approximately 15–20 sensors the size of toasters are deployed per square-mile and typically placed on street lights, rooftops, or on the sides of buildings. When a loud bang-like noise occurs, the sensors triangulate the location based on the time it takes for sound to reach the different sensors. These records are sent to ShotSpotter, where the company's analysts determine whether or not it is gunfire before reporting it to police within 1 minute (Yousef, 2017).

My sample begins on January 13, 2017 and continues to January 14, 2019. Beginning in 2017, four police districts (Englewood, Harrison, Austin and Deering) were outfitted with ShotSpotter technology with plans to expand to Ogden and Gresham districts by July. In October 2017, an expansion to six more districts was announced (Wentworth, Grand Crossing, South Chicago, Calumet, Chicago Lawn and Grand Central) for a total of 12 of the city's 25 districts (and approximately 130 square-miles of area) being covered by ShotSpotter. The first six districts were considered the tier of most violent in Chicago, while the next six were classified as the second most violent tier (Wasney, 2017).

The dataset contains the event date and time, the address (censored at the block level) and the latitude and longitude coordinates.

Table 1 provides descriptive statistics. The number of observations falls when compared to reported crime because of the limited time period and number of districts with ShotSpotter coverage. On average, there are about 1.68 incidents per district-day, and the bulk of those were incidents with multiple shots (0.99 incidents per district-day, or 59%). There are 6.52 rounds fired per day, which is about 3.9 rounds per incident.

2.3.4 Fatal Encounters

Because the CPD TRR data does not contain information about whether or not subjects are fatally injured, I use data from Fatal Encounters to match incidents and determine fatal incidents. Recent research has indicated that these data are quite accurate for fatal incidents (see for example Campbell, Nix, and Maguire, 2018) even though their lack of representativeness of OIS overall is in part a motivation for this paper. These data contain the incident date, location (including longitude and latitude), and cause of death; subject name, age, and gender; as well as links to the news article(s) the information was drawn from. First, I assign the Fatal Encounters incidents in Chicago to the police districts they occured in. Next, I manually match each Fatal Encounters incident to OIS in the CPD TRR data based on the district the OIS occured in, the date of the incident, and subject race and age. Table 1 indicates that, overall, 2.5% of all OIS result in a fatal injury to the subject. These comprises 116 incidents, 101 of which are tied to firearm discharges (this is approximately 14% of all OIS with firearm discharges).

Figure 5 provides several maps of Chicago's police districts highlighting the: number of OIS per 10,000 residents from 2004–2019 (panel a)); share of population that is Black (panel b)); average daily gunshots (detected by ShotSpotter) per 10,000 residents (panel c)); and residents who are average daily reported crime per 10,000 residents(panel d)).[7] The takeaways are that there is variation each in these measures across police districts and that gunshots, OIS, and reported crime tend to be concentrated in districts with higher shares of Black residents.

2.4 Strategy

2.4.1 Injured and Not Injured

To examine the effects of OIS on community crime reporting, I restrict my sample to compare observations 14 days before and 14 days after of an OIS. I then compare crime reporting over a given window in districts where the OIS resulted in a subject injury to reporting over that same window in districts where the OIS did not result in a subject injury. This means that I am comparing crime reporting behavior in two groups of districts that both experienced an OIS on a given day. The only difference between these observations is that the OIS resulted in an injury in some districts and did not in others.

In my main specification I estimate:

$$Y_{dtmy} = \alpha + \beta INJURY_{dtmy} + X'_{dtmy}\lambda + \omega_d + \varphi_m + \mu_y + \nu_{dy} + \varepsilon_{dtmy}$$

where Y_{dtmy} is the reported crime rate in a given district d, on a given day tmy. X_{dtmy} is a vector of controls including officer, subject, and incident characteristics, ω_d, φ_m, μ_y, and ν_{dy} are district, month, year, district-year fixed effects. $INJURY_{dtmy}$ is a dummy that equals one for post-OIS days in districts with an injury and zero for pre-OIS and post-OIS days for districts without an injury. The coefficient of interest is β and provides an estimate of the before and after change in crime reporting in districts that experienced an OIS resulting in an injury compared to district that experienced an OIS that did not result in an injury. As officers operate out of a specific district I cluster the standard errors at the district level to allow errors to be correlated within districts.

I choose the 14 day time horizon because a) if media coverage is mediating this effect (as is suggested in the literature (e.g. Weitzer, 2002, Weitzer and Tuch, 2004)) then these effects may be short-lived, especially as these are much smaller in scale than OIS used in other studies; b) it can take several days for the media to obtain details on an OIS from witnesses, police, and officials and then begin reporting; c)

[7] Population and race data come from the 2010 U.S. Census at the block-level and are matched to police districts.

there is on average an OIS involving firearms in Chicago overall every 8 days in my sample. I later explore the stability of the results to this choice.

2.4.2 Identification Assumption

The identifying assumption required is that the injury status of a subject from a given OIS (D_{dtmy}) is exogenous, conditional on observables (X_{dtmy}). More formally in a potential outcomes framework, this assumption is given as:

$$\boldsymbol{E}[Y_{dt}|X_{dt}, D_{dt} = 1] - \boldsymbol{E}[Y_{dt}|X_{dt}, D_{dt} = 0] = \boldsymbol{E}[Y_{1dt} - Y_{0dt}|X_{dt}]$$

where Y_{dt} is the reported crime, Y_{1dt} is reported crime for a district with an injury-resulting OIS, and Y_{0dt} is reported crime for a district without an injury-resulting OIS.

The intuition of this identification assumption is provided by FBI Special Agent John Huber, "We don't shoot to kill; we shoot to stop." (Lane, 2016). Members of the public often want to know why police officers fire so many times, or why they do not attempt to shoot to disarm, wound, or injure rather the subjects they are pursuing. The answer provided by law enforcement is that in the context where discharging a firearm is an option, officers place themselves, other officers, and other members of the public at risk if they are unable to incapacitate the subject presenting an immediate threat to life or serious injury. If, for example, an officer attempts to shoot a gun out of a subject's hand and misses, not only do they risk harming someone else, the threat posed by the subject is still present. Because officers' first priority in this context is incapacitating the subject, they are trained to target the center mass and central nervous system, which is also likely to result in death. CPD directives prohibit firing warning shots, at subjects who pose only a threat to themselves, fleeing suspects, as well as into buildings, windows, or openings where the subject is not visible. To bring this back to the identification assumption, when an officer discharges their firearm, the intent is always to hit and incapacitate the subject by aiming for their center of mass.[8] Failing to do so is not strategic behavior meant to deter, disarm, or displace, it is the result of officers being unable to hit the subject.

The 1985 US Supreme Court ruling in *Tennessee v. Garner* lead to restrictions in the application of deadly force by police officers. They were no longer able to apply deadly force simply to prevent a subject from escaping, instead the use of deadly force required that the subject pose an immediate danger of death or serious injury to the officer or others. This was followed by the 1989 Supreme Court ruling in *Graham vs. Connor* that now required courts to take the context of a shooting into account, including understand that the decisions of officers happen in a split-second, in order to determine whether or not a shooting is justified. These two rulings backstop the CPD (and all other law enforcement agencies) directives and guidelines on use of deadly force. The CPD's directive indicates that officers may only use force only as a last resort to protect against "an imminent threat to life or to prevent great bodily harm to the member or another person" (CPD, 2017).

This identification assumption requires one-to-one matching between subjects and officers, however there are some incidents with multiple officers (10.7% of incidents have multiple officers) and/or multiple subjects (4.8% of incidents have multiple subjects). I restrict my sample to taking the characteristics of the oldest officer in

[8]Technically speaking, TASER discharges are meant to target the "lower-center of mass" which provides better incapacitation and a directive to this effect was issued by the TASER company in 2009 (TASER, 2009).

OIS with multiple officers, and the youngeset subject in OIS with multiple subjects. I choose the oldest officer as they should be in charge of the sitaution and most likely to issue commands. I choose the youngest subject as they may be most likely to behave unpredictably, resulting in the escalation of the incident. Regardless, I explore the sensitivity of my estimates to these decisions in Section 2.5.

2.4.3 Predicting Injury Status

Table 2 presents the unconditional balance for covariates split by subject injury status. Though the identification strategy is about conditioning on these observables, it is interesting to first examine how they breakdown between injured and uninjured subjects. First, we can see that CPD members involved in OIS where subjects are injured are more often male, less experienced, more likely to be an officer, more likely to be injured, more likely to discharge a hand gun (and consequently less likely to discharge a TASER). Appendix Table A1 provides the same breakdown restricting to only OIS where officers discharge a firearm and indicates that only an officer's injury status and likelihood of discharging a hand gun (in this case the alternative weapon is a long gun) are statistically different. This suggests that much of what is captured here is the difference between officers who discharge firearms and who discharge TASERs. The main results I present are for firearms only, though when pooling the two, I control for discharged weapon type.

Next, we can see that subjects who are injured are less likely to: be Black and more likely to be White or Hispanic; be armed; have a drug, weapon, or assault charge, have a hand gun. When examining only firearm discharges, differences disappear between subject race, and weapons and assault charges but remain for the others (these OIS that result in injury have subjects who are more likely to have a domestic charge).

Lastly, OIS that result in subject injury have: fewer counts of active resistance; more counts of passive resistance; subjects more likely to use deadly force; officers who fire first; artificial lighting conditions and happen indoors. For OIS with firearm discharges differences disappear for counts of active resistance and whether or not the officer fired first but remain for the others (these OIS that result in injury are also more likely to happen in daylight).

Overall, even unconditionally there is some balance between OIS with injured and uninjured subjects, especially among OIS with firearm discharges. Regardless, I control for all of these covariates in my analysis.

Next, to provide suggestive evidence in support of this identification assumption, I estimate a linear probability model where the dependent variable is whether or not a subject is injured in a given OIS. I include the full suite of officer, subject, and incident characteristics. The purpose of this exercise is to examine how predictive of injury status these covariates are after accounting for the full set. If it is difficult to predict injury status, we should expect to see few statistically significant results and coefficients that are very close to 0. In fact, this is precisely what I find. Figure 8 presents the coefficients and standard errors from this exercise for officer level characteristics, while Figures 9 and 10 do the same for subject and incident characteristics, respectively. Solid circles indicate that the estimate is statistically significant at the 5% level while hollow circles indicate the the estimate is not statistically significant at the 5% level. The type of weapon used by a subject, the weather and lighting conditions, whether an OIS is indoors, officer experience, TRR history, and age are all are not predictive of subject injury status. Only a few of these characteristics are statistically significant–female, Black, or Hispanic officers are all less likely to injure an subject, conditional on discharging their weapon, as are officers not in uniform; incidents involving a weapons charge are less likely to result in an injury while those with an assault charge are more likely to result in an injury. Interestingly, incidents where neither party fired first are more likely to result in an injury. However, even those characteristics that are statistically significant are clustered around zero, indicating that no one characteristic explains subject injury status well after accounting for the others.

Overall, even unconditionally there is some balance between OIS with injured and uninjured subjects, especially among OIS with firearm discharges, and many of these covariates do not appear to be good predictors of subject injury status. Regardless, I control for all of these covariates in my analysis.

2.5 Results

In this section I first present event study results, then the comparisons of crime reporting in districts with OIS resulting in injury, and finally results related to mechanisms and effect heterogenity.

2.5.1 Injury v.s. Non-Injury OIS

Table 3 presents my main results from Equation (2.4.1). The dependent variable is ln(Reported Crime + 0.1). Only district-day observations up to 14 days before and after an OIS are included. The pattern that reveals itself is that OIS with an injured subject reduce reported crime more than OIS with uninjured subjects in the days following the attacks. Column 1 includes officer controls and I find that OIS with subject injured reduce reported crime in the 14 days following by about 3.7%. Columns 2 and 3 add subject and incident controls, respectively, and have nearly identical point estimates as column 1. Column 4 brings in month fixed-effects and column 5 adds district-year fixed effects, both of which shrink the estimated effect size (down to 3.1% and 2.8% respectively). All columns include district and year fixed effects. Column 5, with its district specific annual trends in crime is my preferred specification. In Appendix Table A2, I present these estimates with 95% confidence intervals adjusted for the small number of clusters in my sample (22 police districts) using the fast wild bootstrap with Rademacher weights (999 iterations). These adjusted confidence intervals are presented in square brackets and located under the cluster-robust standard errors. The estimates are robust to these small cluster adjustments.

Appendix Figure A1 provides an event study type plot of ln(Reported Crime + 0.1) in districts with OIS where officer discharged firearms resulting in injury at daily intervals for the 21 days before and after the OIS. The full suite of officer, subject, and incident controls as well as year, month, district, and district-year fixed effects are included. What emerges is a clear shift downwards in point estimates on reported crime after an OIS; confidence intervals that exclude zero for many of these daily level estimates; and suggestive evidence that the effect is not persistent beyond a 14 day period. Even though this is not a difference-in-differences research design, divergent pre-trends may still call into question whether or not a subject's injury status as good as randomly assigned, conditional on these controls. That the figure does not indicate pre-trend issues provides some reassurance that the use of districts with OIS that do not result in subject injury as the control group is appropriate.

Appendix Table A3 presents the analogous results including TASER incidents. Recall that in this sample 85% of OIS are TASER incidents and the remaining 15% are firearms. When TASER events are included, the estimated effect is not statistically significant and is much smaller in magnitude with quite tight confidence intervals. When taken with the results above, this suggests that while reported crime does not respond to OIS where officers discharge TASERs, it does respond to incidents where officers discharge firearms. This may mean that TASER incidents are not as salient to the public and therefore less likely to alter reporting behavior, or this may reflect a lack of deterrence effects on crime by TASERS.

In Appendix Tables A4–A6, I test the sensitivity of these results to the alternative choices of officer and subject characteristics. Appendix Table A4 provides estimates when characteristics are instead drawn from the oldest officer and the oldest subject; A5 does the same for the youngest officer and the oldest subject; and A6 for the youngest officer and the youngest subject. The point estimates range from 3.8%–2.7% across specifications–nearly identical to the 3.7%–2.8% range found in Table 3. These results provide reassurance that my findings are not being driven by the selection of officer and subject characteristics required by my identification strategy.

In Appendix Table A7, I explore the sensitivity of my findings to alternative choices of pre- and post-OIS periods. The baseline for equation 2.4.1 is 14 days before and 14 days after the OIS. Column 1 increases the pre-period to 21 days while maintaining the post period of 14 days. Column 2 returns the pre-period to the baseline 14 days and shrinks the post OIS period to 10 days. Column 3 estimates with a pre-period of 10 days and a post-period of 14 days, while column 4 has a pre-period of 14 days and a post-period of 21 days. Column 5 increases repeats the exercise setting pre- and post-periods both to 21 days. All columns include the full suite of controls and fixed effects (equivalent to column 5 of Table 3). The estimated effects range from 1.9–3.7% compared to the 2.8% in my main specification and are all statistically significant at the 1% level (except column 5 which is statistically significant at the 5% level). These findings suggest that moving the pre- and post-OIS interval backward and forward from the date of the OIS has no effect on the main conclusions of this paper.

Following the discussion in Section 2.2, I test whether or not these reductions in reported crime following OIS with injured subjects are limited to serious, violent, or less serious crimes. Table 4 provides provides estimates of column 5 in Table 3 but switches the dependent variables from a measure of all reported crimes to these subsets. Column 1 simply repeats the estimates for all reported crimes. The dependent variable in column 2 is more serious crime–only reported Part I Index Crimes (homicide, rape, robbery, aggravated assault, burglary, motor vehicle theft, arson). In column 3 the dependent variable is violent crime (homicide, rape, robbery, and aggravated assault). Lastly, column 4 has a dependent variable measuring reported less serious offenses (things like narcotics, weapons, and liquor violations or simple assault). These dependent variables have all been log transformed (log (x+ 0.1)) to aid with interpretation and to address zeros that occur in these smaller subsets for all reported crime. When splitting reported crime up this way, the point estimates are all negative and roughly the same magnitude as column 1 (all reported crime), staying around 2.2%–3.3% reductions in reported crime. However, only column 4 measuring the effects of OIS with injured subjects on reported less serious crime is statistically significant at conventional levels. As noted in Table 1, about 60% of reported crime is less serious, making inference on serious and violent crime less precise. As such, I cannot rule out equivalence between the estimated effects on the different types of crime.

2.5.2 Channels and Heterogeneity

Because changes in both reporting behavior and criminal behavior are consistent with the data, I turn to ShotSpotter in an attempt to disentangle them.

Table 5 repeats the analysis in Table 3 but instead uses the natural log of the count of ShotSpotter incidents as the dependent variable. The results suggests that it is unclear that there is a difference in detected ShotSpotter incidents in districts with an OIS where officers discharge firearms and injure someone compared to districts without such an injury. The point estimate is negative and, given that ShotSpotter data exists for only 13 districts, the standard errors are quite large and therefore these estimates are quite imprecise.

Appendix Table A8 provides estimates of Equation (2.4.1) while incorporating ShotSpotter data. In column 1, I re-estimate the results presented in Table 3 but restrict the sample to only the districts with ShotSpotter technology over the entire sample period. I do this to check that any results obtained when controlling for ShotSpotter events are not simply a result of these districts having different behavior than the other districts. In fact, the estimated effect for these districts is directly in line with that from Table 3. In column 2, I repeat this exercise but now restrict the sample to only the period where I have ShotSpotter data (2017-2019). Lastly, in column 3 I control for the number of ShotSpotter events on a given district-day.

The results from columns 2 and 3 highlight several important points. First, turning to the last row of column 3, ShotSpotter incidents are positively correlated with reported crime. This is reassuring in that crime and reported crime move together and in the same direction. Second, and perhaps most notably, the previous findings of negative and statistically significant point estimates disappear. Looking at the size of the sample for these columns, this might be unsurprising–the standard errors on the coefficient of interest are ten times larger than those in previous estimates. Lastly, even after controlling for the number of ShotSpotter events, the point estimate and standard errors of the coefficient of interest are unchanged. So, while this sample is very small and we cannot rule out relatively large positive or negative effects of OIS on reported crime, these results do provide suggestive evidence that accounting for some measure of underlying crime does not dramatically alter the estimates of crime reporting behavior.

Table 6 examines whether or not there are differential effects of these OIS. Column 1 repeats the results from column 5 of Table 3, my preferred specification. Row 1 presents the baseline interaction and coefficient of interest from Equation (2.4.1). Column 2 adds a triple interaction term, looking for differential effects for OIS resulting in an injury when a subject is unarmed compared to when a subject is armed. I find that reported crime falls by an additional 4.8% following an OIS in districts where a subject is injured and unarmed than in districts with an OIS where a subject is injured but unarmed.

Column 3 introduces an interaction for whether or not the subject was fatally injured by the OIS. Interestingly, there does not appear to be additional reduction in reporting when the subject is killed. If the decline in reporting rate is driven by deterrence-induced reductions in crime a la the Becker model, then we should expect to see further reductions in reported crime when the subject is fatally injured than when simply injured as that represents a much larger shock to the perceived costs of engaging in crime. This is especially true because the fatal encounters data is drawn from media sources, meaning that salience should be high for these OIS. The results from column 3 provide some suggestive evidence in favor of the crime reporting behavior channel rather than the reduction in underlying crime channel.

Column 4 adds a similar interaction term to explore differing effects when the subject is Black. While reported crime still falls after an OIS with an injury compared to an OIS without, there does not appear to be different effects in reported crime when the subject is Black compared to when the subject is not Black. Column 5 repeats the exercise for OIS in majority Black districts and again I do not find differential effects in reported crime following an OIS resulting in an injury. Column 6 examines whether or not OIS resulting in injury might reduce reported crime more when the officer and subject do not share a race. Specifically, it looks at White officers and Black subjects and the results do not indicate differential effects on reported crime. I find that there are no differential impacts on reported crime after an OIS when the officer is White and the subject is Black.

The results from columns 4–6 are somewhat surprising given evidence in this literature that white officers are much more likely to use force against Black subjects and in minority neighborhoods (see for example Ba et al., 2021; Hoekstra and Sloan, 2020) and that the discussion in popular media often centres the contrasting race of the subject and officer in coverage. Combining these results with those from column 2 suggests that procedural justice matters more in determining police legitimacy rather than distributive justice and that police firing on unarmed subjects is perceived as procedurally unjust, regardless of the CPD's internal policy.

Appendix Tables A18, A19, A20 repeat this analysis for each of the different types of crime. Overall, the same pattern of no apparent effects along racial dimensions but negative effects for unarmed subjects appears. For serious crime, these race-based coefficients are all very small, though somewhat imprecisely estimated. The exception is for OIS with fatally injured subjects which indicates an increase in reported crime, statistically significant at the 5% level. This coefficient is also positive for less serious crimes though is not statistically significant and is less precisely estimated. However, for less serious crime the estimated effect for OIS with unarmed subjects who are injured is statistically significant at the 5% level, quite large (an additional 6.5% reduction in reported less serious crimes), and we can rule out equivalence with that of serious crimes. This is consistent with the notion that we might expect to see larger reductions in reported crime among crimes where witnesses or bystanders have more discretion in reporting.

2.6 Conclusion

In this paper I estimate the impacts of OIS on crime reporting in Chicago by exploiting the natural randomness in subject injury status. I find that OIS that result in an injury reduce reported crime by about 2.8% in comparison to OIS that do not result in injury. Furthermore, I do not find differential effects on crime reporting when the subject is Black, when the officer-subject pair is White-Black, nor when the OIS occurs in a majority Black district. However, I do find that when a subject is unarmed the decline in reported crime is even larger (an additional 4.8% reduction). The lack of differential effects along racial dimensions suggests that citizens' perceptions of police legitimacy are not driven by notions of distributive justice and when paired with the results about unarmed subjects indicate that it is likely notions of procedural justice that more strongly influence perceptions of police legitimacy.

These findings are in line with those of the qualitative literature in sociology and criminology, that a procedural justice model is how citizens evaluate police legitimacy. I also provide mild suggestive evidence that accounting for a measure of crime does not dramatically shift the estimated effect but when paired with the a lack of differential effects on reporting after OIS with a fatally injured subject suggest that these changes in reported crime are unlikely to simply reflect lower underlying criminal activity.

Overall, these findings are relevant for understanding the consequences of OIS on the efficiency and effectiveness of public safety provision more broadly and police services more specifically. Most notably, these reductions in citizen assistance may now be partly responsible for the reductions in police clearance rates which have previously been characterized as resulting from lower proactive policing activities. For example, if police departments sought to combat lower clearance rates by maintaining levels of proactive policing, my results suggest that clearance rates may still suffer.

This area is still in need of further research. My results are unable to say anything about the other channels through which reductions in police legitimacy may impact public safety provision and policing such as non-cooperation with police during ongoing investigations (e.g. refusal to pick suspects out of a lineup, provide testimony, or give witness statements) or reliance on a code of the street to deliver justice. In order to address these channels more detailed data of police investigations or criminal activity is needed. Additionally, more work is needed to examine the direct and indirect effects of OIS and police legitimacy on other aspects of the criminal justice system and police efficiency.

2.7 Figures

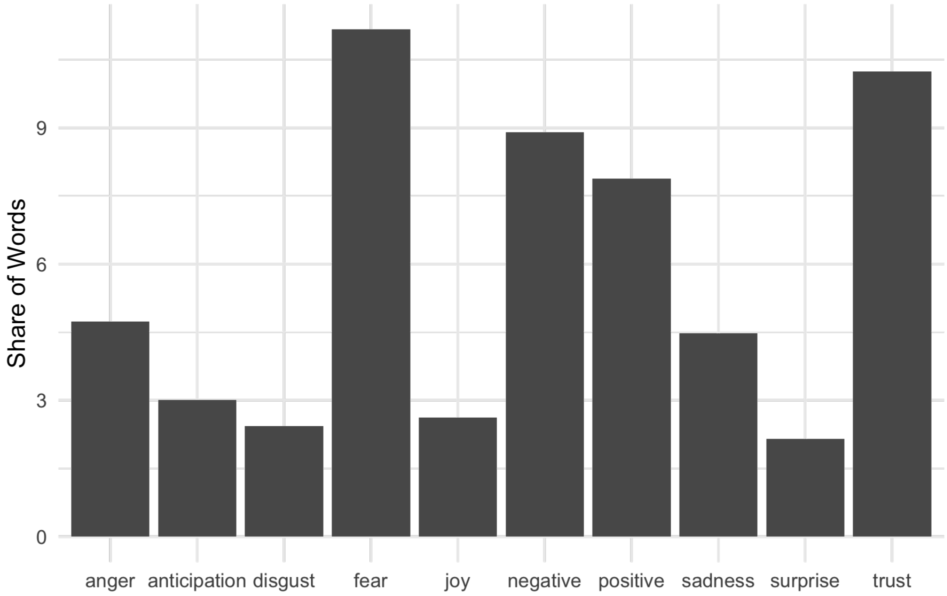


FIGURE 1: Share of words per tweet classified as positive, negative, or each emotion. Time period is July 1—31 2017. Tweets selected to include “cops”, “cop”, or “police” and are rstricted to the United States.

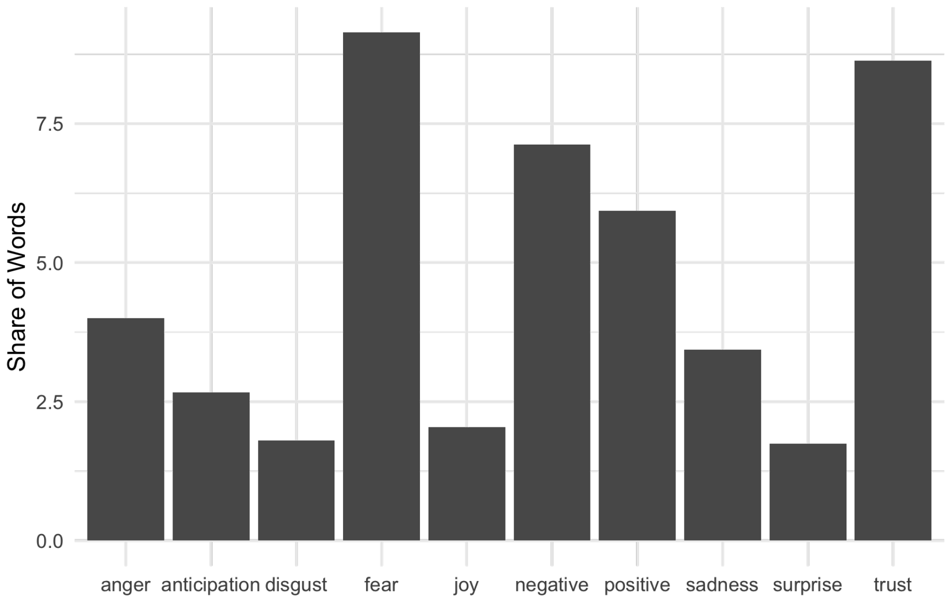


FIGURE 2: Share of words per tweet classified as positive, negative, or each emotion. Time period is March 1—31 2018. Tweets selected to include “cops”, “cop”, or “police” and are rstricted to the United States.

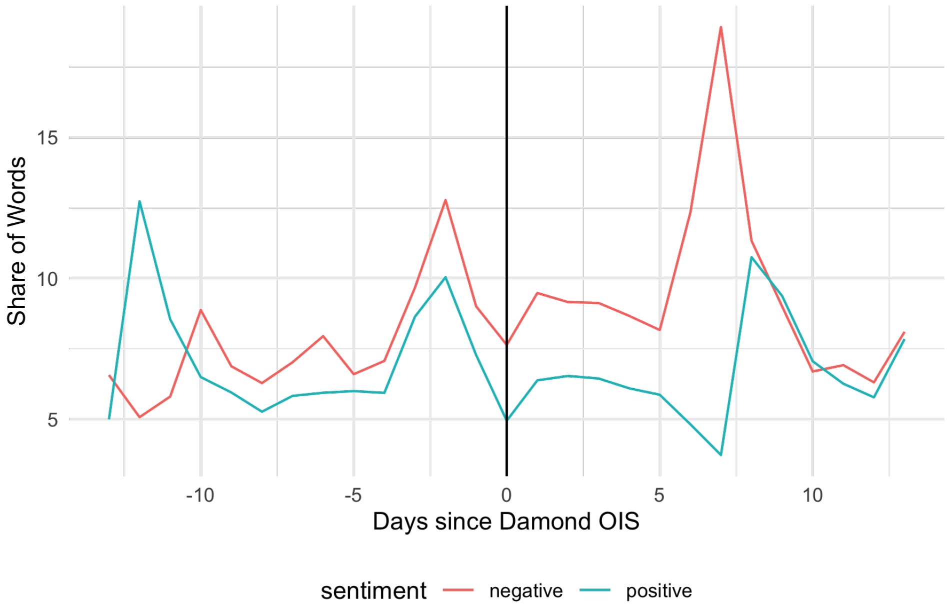


FIGURE 3: Share of words per tweet classified as positive or negative. Time period is July 1—31 2017. Tweets selected to include "cops", "cop", or "police" and are rstricted to the United States.

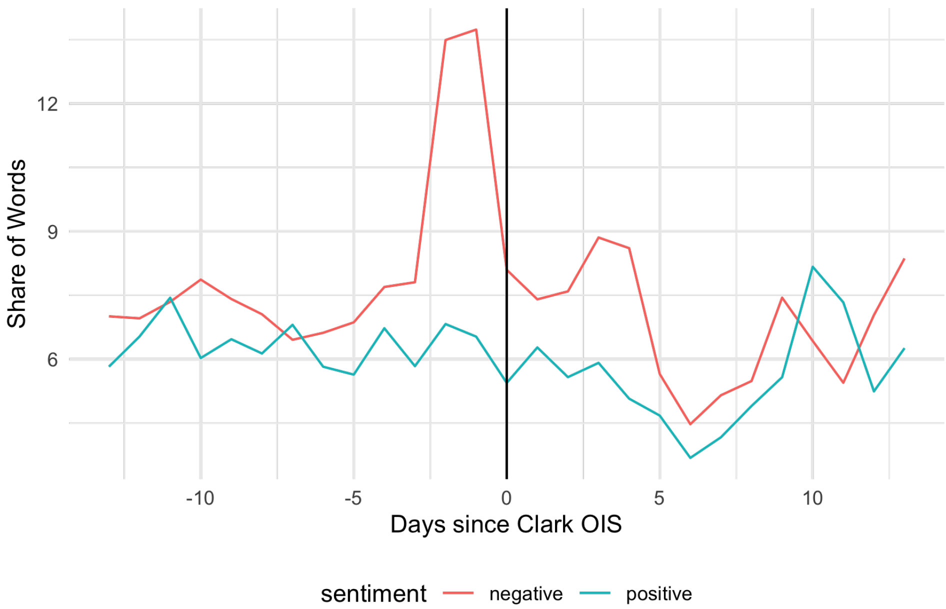


FIGURE 4: Share of words per tweet classified as positive or negative. Time period is March 1—31 2018. Tweets selected to include "cops", "cop", or "police" and are rstricted to the United States.

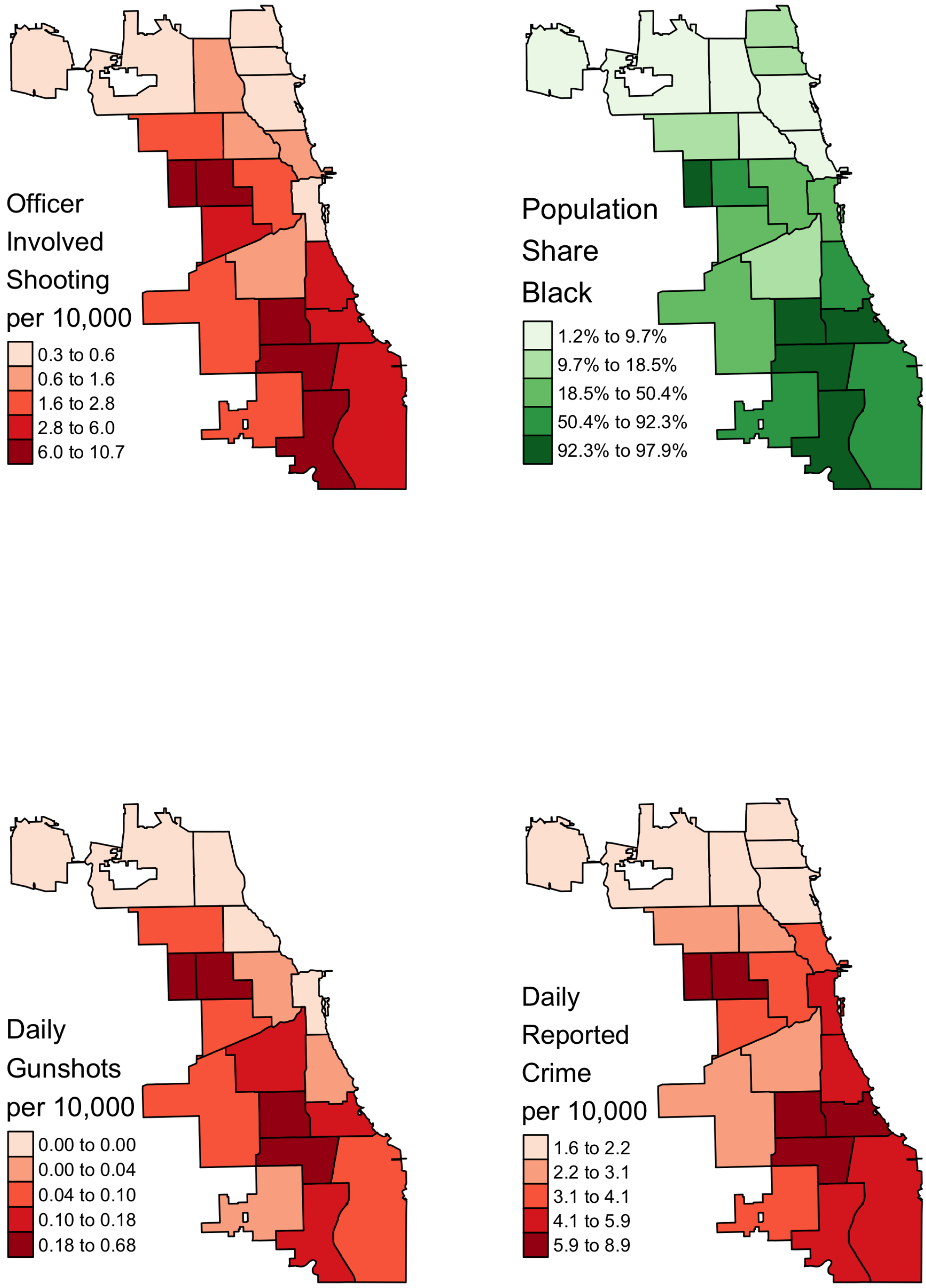


FIGURE 5: In each panel, legend breaks correspond to quartiles. Panel a) plots officer-involved shootings per 10,000 residents from 2004–2019. Panel b) plots each districts share of residents who are Black. Panel c) plots ShotSpotter incidents per 10,000 residents from 2017–2019. Panel d) plots mean reported crime per 10,000 residents from 2004–2019.

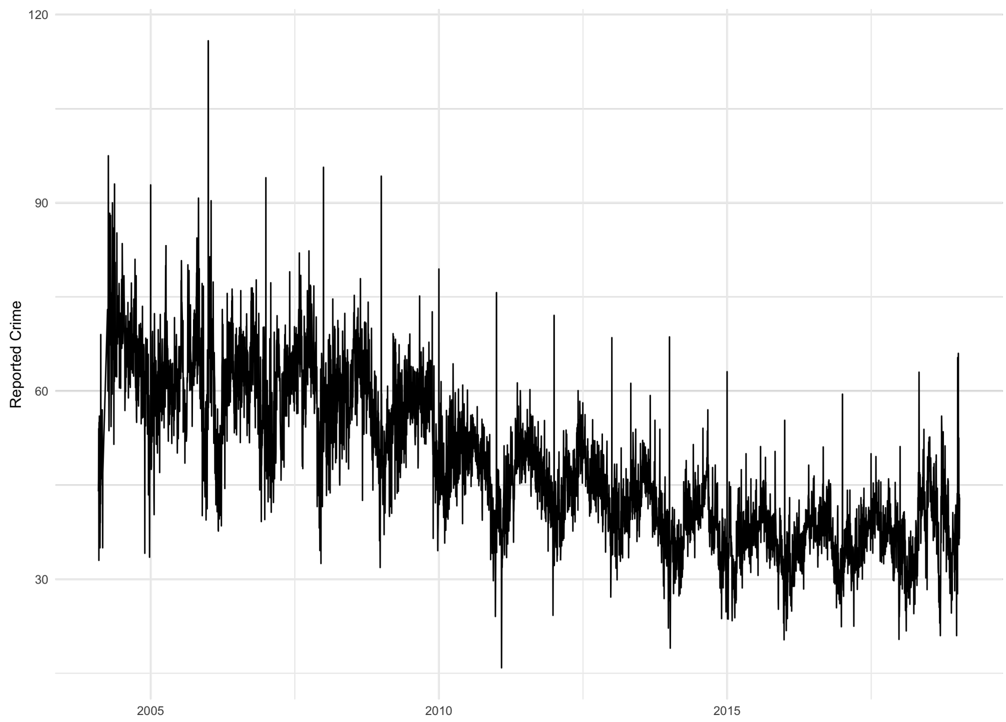


FIGURE 6: This figure plots the daily average reported crimes per district over the sample period (2004–2019).

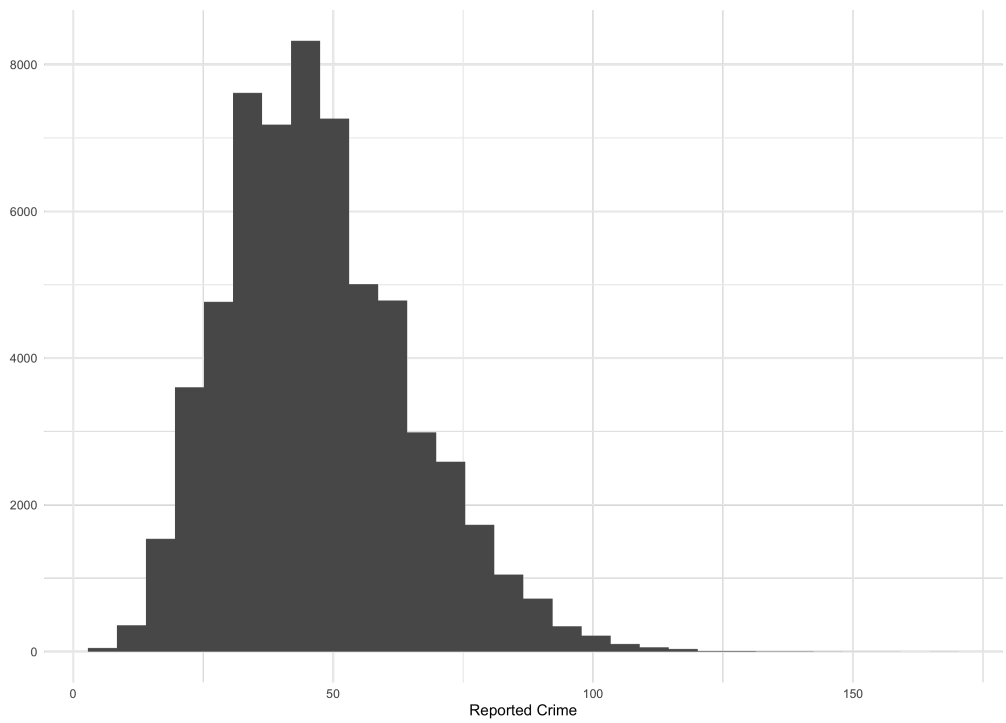


FIGURE 7: This figure plots a histogram of reported crimes over the sample period (2004–2019).

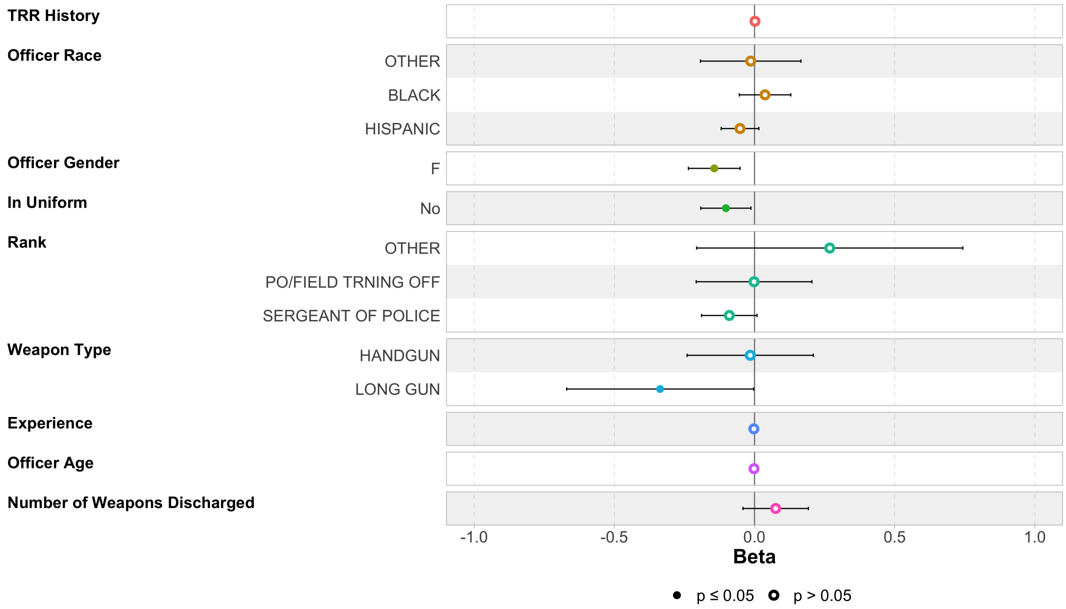


FIGURE 8: This figure plots estimated coefficients from regressing subject injury on all officer, subject, and incident characteristics. Only coefficients from officer variables are presented in this figure. Full circles are statistically significant at the 5% level, hollow circles are not statistically significant at the 5% level.

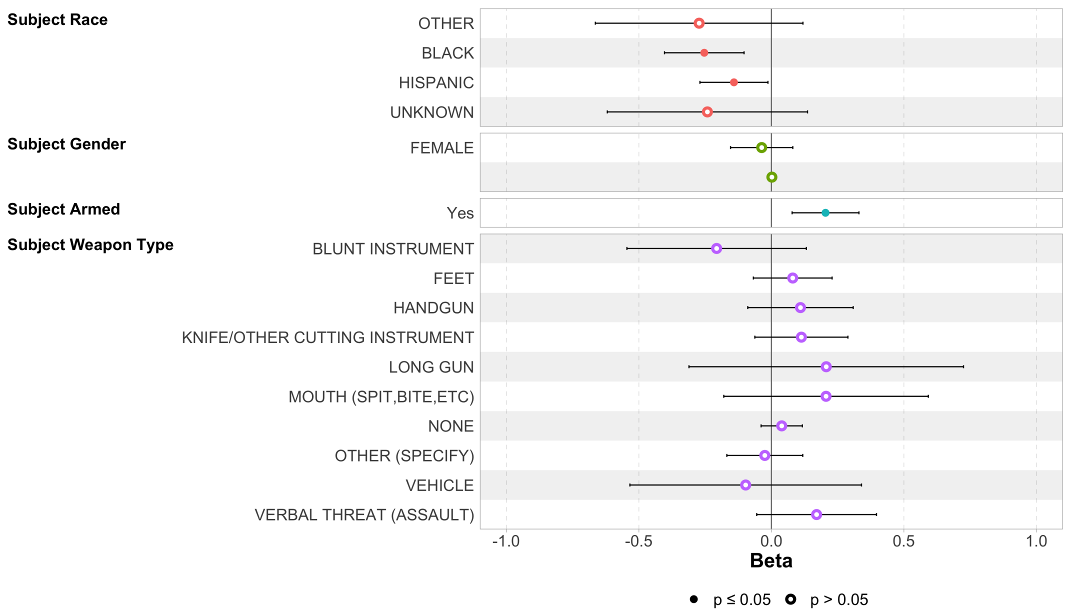


FIGURE 9: This figure plots estimated coefficients from regressing subject injury on all officer, subject, and incident characteristics. Only coefficients from subject variables are presented in this figure. Full circles are statistically significant at the 5% level, hollow circles are not statistically significant at the 5% level.

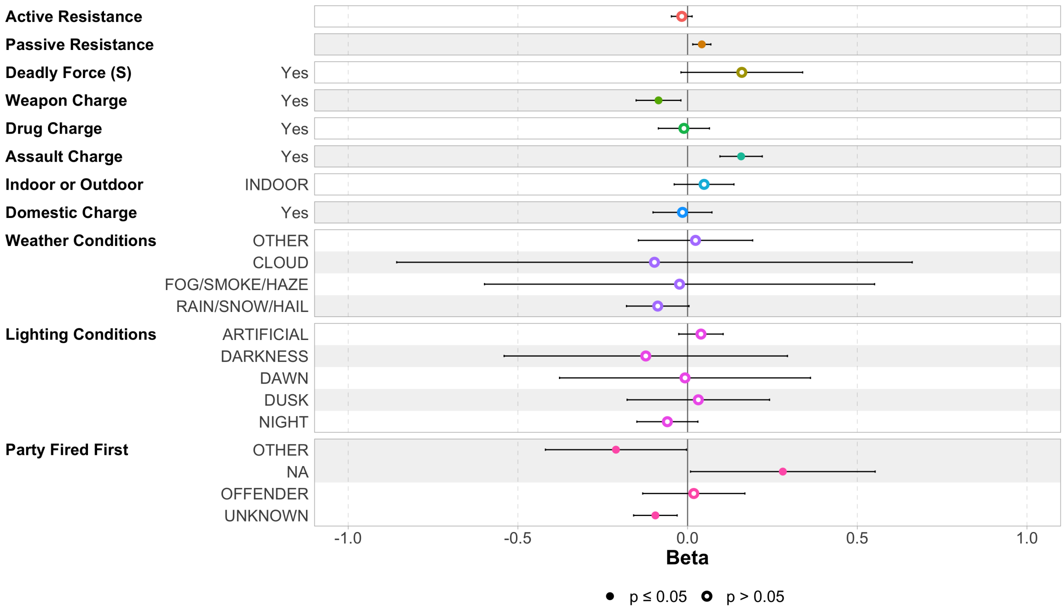


FIGURE 10: This figure plots estimated coefficients from regressing subject injury on all officer, subject, and incident characteristics. Only coefficients from incident variables are presented in this figure. Full circles are statistically significant at the 5% level, hollow circles are not statistically significant at the 5% level.

2.8 Tables

TABLE 1: Descriptive Statistics

	Mean	Median	P5	P95	N
Reported Crime					
Reported Crime	47.19	45.00	22.00	80.00	60,365
Reported Less Serious Crime (Part II Index)	28.67	27.00	11.00	52.00	60,365
Reported More Serious Crime (Part I Index)	18.52	18.00	7.00	33.00	60,365
Reported Violent Crime (Part I Index)	3.01	3.00	0.00	7.00	60,365
ShotSpotter Incidents					
Multiple Shot ShotSpotter Incidents	0.99	0.00	0.00	5.00	5,274
ShotSpotter Incidents	1.68	0.00	0.00	7.35	5,274
ShotSpotter Rounds	6.52	0.00	0.00	32.00	5,274
Single Shot ShotSpotter Incidents	0.51	0.00	0.00	3.00	5,274
Officer Characteristics					
Age	37.79	37.00	27.00	52.00	4,654
Black	0.18	0.00	0.00	1.00	4,654
Experience (Yrs)	10.34	9.56	1.34	23.09	4,654
Hispanic	0.26	0.00	0.00	1.00	4,654
In Uniform	0.76	1.00	0.00	1.00	4,654
Male	0.89	1.00	0.00	1.00	4,654
Number of weapons discharged	1.06	1.00	1.00	2.00	4,654
Officer fired Handgun	0.15	0.00	0.00	1.00	4,654
Officer fired Long gun	0.00	0.00	0.00	0.00	4,654
Officer fired TASER	0.85	1.00	0.00	1.00	4,654
Prior Incidents	9.36	6.00	0.00	29.00	4,654
Rank: Officer	0.74	1.00	0.00	1.00	4,654
White	0.52	1.00	0.00	1.00	4,654
Subject Characteristics					
Age	30.15	27.00	19.00	50.00	4,654
Armed	0.24	0.00	0.00	1.00	4,654
Black	0.77	1.00	0.00	1.00	4,654
Hispanic	0.14	0.00	0.00	1.00	4,654
Injured	0.45	0.00	0.00	1.00	4,654
Male	0.93	1.00	0.00	1.00	4,654
Weapon: Fists	0.13	0.00	0.00	1.00	4,654
Weapon: Handgun	0.11	0.00	0.00	1.00	4,654
Weapon: Knife	0.00	0.00	0.00	0.00	4,654
Weapon: Long Gun	0.00	0.00	0.00	0.00	4,654
Weapon: None	0.54	1.00	0.00	1.00	4,654
White	0.07	0.00	0.00	1.00	4,654
Subject Fatally Injured	0.02	0.00	0.00	0.00	4,655
Incident Characteristics					
Active resistance	2.75	3.00	1.00	5.00	4,654
Artificial light	0.55	1.00	0.00	1.00	4,654
Assault charge	0.63	1.00	0.00	1.00	4,654
Clear weather	0.88	1.00	0.00	1.00	4,654
Cloudy weather	0.00	0.00	0.00	0.00	4,654
Darkness	0.01	0.00	0.00	0.00	4,654
Dawn	0.00	0.00	0.00	0.00	4,654
Daylight	0.31	0.00	0.00	1.00	4,654
Deadly force (S)	0.17	0.00	0.00	1.00	4,654
Domestic charge	0.14	0.00	0.00	1.00	4,654
Drug charge	0.19	0.00	0.00	1.00	4,654
Fog/Haze/Smoke weather	0.00	0.00	0.00	0.00	4,654
Indoor	0.25	0.00	0.00	1.00	4,654
Officer fired first	0.70	1.00	0.00	1.00	4,654
Passive resistance	3.41	3.00	1.00	6.00	4,654
Rain/Snow/Hail weather	0.08	0.00	0.00	1.00	4,654
Weapon charge	0.13	0.00	0.00	1.00	4,654

Note: Time period is 2004–2019 (ShotSpotter 2017–2019). Including TASER events there are 60,365 district-day observations. ShotSpotter data's reduced time period means there are only 5,724 district-day observations. Including TASER events there are 4,654 officer-involved shootings in the sample. "Prior Incidents" is a variable that equals the count of previous tactical response report incidents for a given officer at the time of the shooting. "Passive resistance" are counts of passive resistance in an incident (e.g. failure to comply with a verbal command). " Active resistance" are counts of active resistance in an incident (e.g. physically resisting arrest). "Deadly force (S)" is a binary variable that takes 1 if a subject used deadly force against an officer.

TABLE 2: Unconditional Covariate Balance by Subject Injury Status

	Subject Uninjured			Subject Injured			
Variable	N	Mean	SD	N	Mean	SD	Test
Officer Characteristics							
Officer Black	2568	0.172	0.378	2086	0.181	0.385	F=0.655
Officer White	2568	0.519	0.5	2086	0.529	0.499	F=0.441
Officer Hispanic	2568	0.273	0.446	2086	0.253	0.435	F=2.336
Officer Male	2568	0.878	0.328	2086	0.905	0.294	F=8.489***
Officer Age	2568	37.913	7.676	2086	37.647	7.573	F=1.399
Experience (Yrs)	2568	10.532	6.794	2086	10.103	6.779	F=4.594**
Prior Incidents	2568	9.27	9.576	2086	9.471	10.789	F=0.452
In Uniform	2568	0.755	0.43	2086	0.774	0.418	F=2.245
Rank: Officer	2568	0.725	0.447	2086	0.76	0.427	F=7.611***
Officer Injured	2568	0.084	0.277	2086	0.131	0.338	F=27.923***
Officer Fired Handgun	2568	0.111	0.315	2086	0.187	0.39	F=53.56***
Officer Fired TASER	2568	0.882	0.322	2086	0.81	0.393	F=48.156***
Number of weapons discharged	2568	1.055	0.245	2086	1.066	0.277	F=2.121
Subject Characteristics							
Subject Black	2568	0.797	0.402	2086	0.746	0.435	F=17.311***
Subject White	2568	0.055	0.227	2086	0.082	0.275	F=14.402***
Subject Hispanic	2568	0.127	0.333	2086	0.153	0.36	F=6.541**
Subject Male	2568	0.934	0.249	2086	0.936	0.244	F=0.113
Subject Age	2568	29.948	10.263	2086	30.388	10.298	F=2.114
Subject Armed	2568	0.191	0.393	2086	0.297	0.457	F=72.266***
Domestic charge	2568	0.134	0.341	2086	0.146	0.353	F=1.349
Drug charge	2568	0.204	0.403	2086	0.17	0.376	F=8.63***
Assault charge	2568	0.592	0.492	2086	0.685	0.465	F=42.586***
Weapon charge	2568	0.122	0.327	2086	0.146	0.353	F=5.697**
Weapon: None	2568	0.584	0.493	2086	0.484	0.5	F=46.787***
Weapon: Fists	2568	0.129	0.336	2086	0.128	0.334	F=0.032
Weapon: Handgun	2568	0.083	0.275	2086	0.135	0.342	F=33.836***
Weapon: Long Gun	2568	0.002	0.044	2086	0.004	0.066	F=2.151
Incident Characteristics							
Active resistance	2568	2.815	1.362	2086	2.675	1.497	F=11.122***
Passive resistance	2568	3.377	1.291	2086	3.445	1.3	F=3.169*
Deadly force (S)	2568	0.132	0.339	2086	0.219	0.414	F=62.344***
Officer fired first	2568	0.682	0.466	2086	0.714	0.452	F=5.596**
Daylight	2568	0.318	0.466	2086	0.296	0.457	F=2.701
Artificial light	2568	0.527	0.499	2086	0.572	0.495	F=9.276***
Darkness	2568	0.005	0.074	2086	0.006	0.076	F=0.019
Dawn	2568	0.005	0.068	2086	0.005	0.069	F=0.004
Clear weather	2568	0.873	0.333	2086	0.879	0.327	F=0.294
Cloudy weather	2568	0.002	0.044	2086	0.002	0.049	F=0.109
Fog/Haze/Smoke weather	2568	0.003	0.056	2086	0.003	0.058	F=0.021
Rain/Snow/Hail weather	2568	0.085	0.279	2086	0.074	0.262	F=1.916
Indoor	2568	0.234	0.423	2086	0.261	0.439	F=4.443**

Note: Time period is 2004–2019. "Prior Incidents" is a variable that equals the count of previous tactical response report incidents for a given officer at the time of the shooting. "Passive resistance" are counts of passive resistance in an incident (e.g. failure to comply with a verbal command). " Active resistance" are counts of active resistance in an incident (e.g. physically resisting arrest). "Deadly force (S)" is a binary variable that takes 1 if a subject used deadly force against an officer. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 3: Effect of OIS involving firearms on crime reporting

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.023*	0.018	0.018	0.011	0.014
	(0.013)	(0.013)	(0.012)	(0.011)	(0.011)
OIS	0.030***	0.029***	0.027***	0.023***	0.020**
	(0.010)	(0.009)	(0.008)	(0.006)	(0.007)
Subject Injured x OIS	-0.037**	-0.036**	-0.037**	-0.031***	-0.028***
	(0.015)	(0.015)	(0.014)	(0.010)	(0.010)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 4: Effect of OIS involving firearms on types of reported crime

	All	Serious	Violent	Less Serious
	(1)	(2)	(3)	(4)
Subject Injured	0.014	-0.000	0.017	0.023
	(0.011)	(0.014)	(0.039)	(0.014)
OIS	0.020**	0.013	0.046	0.024**
	(0.007)	(0.011)	(0.034)	(0.011)
Subject Injured x OIS	-0.028***	-0.022	-0.031	-0.033***
	(0.010)	(0.017)	(0.047)	(0.011)
Observations	10,276	10,276	10,276	10,276
Clusters	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193

Note: Dependent variable in column 1 is ln(Reported Crime + 0.1), in column 2 is ln(Reported Serious Crime + 0.1), in column 3 is ln(Reported Violent Crime + 0.1), and in column 4 is ln(Reported Less Serious Crime + 0.1). Serious crimes are UCR Part I index crimes. Violent crimes are a subset of UCR Part I index crimes: homicide, rape, robbery, and aggravated assault. Less serious crimes are UCR Part II index crimes. All columns include district and year fixed effects. All columns include officer, subject, and incident contols as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 5: Effect of OIS involving firearms on ShotSpotter incidents

	ln(ShotSpotter + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.615	0.026	-7.658***	0.352	0.995
	(1.036)	(0.197)	(0.180)	(3,305.982)	(9,399.729)
OIS	0.272	0.268	0.268	0.311	0.311
	(0.165)	(0.167)	(0.167)	(0.186)	(0.190)
Subject Injured x OIS	-0.064	-0.123	-0.123	-0.222	-0.222
	(0.196)	(0.186)	(0.187)	(0.216)	(0.220)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	413	413	413	413	413
Clusters	13	13	13	13	13
Mean (dep. var)	1.684	1.684	1.684	1.684	1.684

Note: Dependent variable is ln(ShotSpotter + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 6: Heterogeneous effect of OIS involving firearms on reported crime

	ln(Reported Crime + 0.1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subject Injured x OIS	-0.028***	-0.021*	-0.025**	-0.047*	-0.038**	-0.034*
	(0.010)	(0.011)	(0.010)	(0.023)	(0.018)	(0.018)
Subject Injured x OIS x Subject Unarmed		-0.048**				
		(0.020)				
Subject Injured x OIS x Subject Dead			0.060			
			(0.058)			
Subject Injured x OIS x Black Subject				0.027		
				(0.025)		
Subject Injured x OIS x Black District					0.017	
					(0.021)	
Subject Injured x OIS x Different Races						0.014
						(0.022)
Observations	10,276	10,276	10,277	10,276	10,276	10,276
Clusters	22	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include officer, subject, and incident contols as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

2.9 Appendices

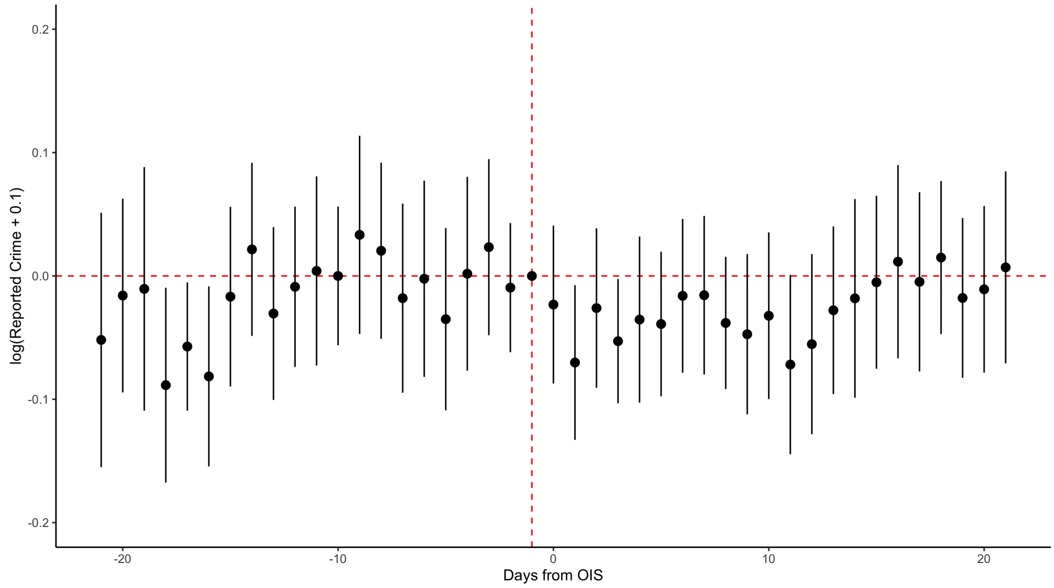


FIGURE A1: This figure plots ln(Reported Crime + 0.1) in districts with OIS where officer discharged firearms resulting in injury at daily intervals for the 21 days before and after the OIS. The full suite of officer, subject, and incident controls as well as year, month, district, and district-year fixed effects are included.

TABLE A1: Unconditional Covariate Balance by Subject Injury Status, OIS with firearms only

	Subject Uninjured			Subject Injured			
Variable	N	Mean	SD	N	Mean	SD	Test
Officer Characteristics							
Officer Black	302	0.228	0.421	397	0.249	0.433	F=0.409
Officer White	302	0.444	0.498	397	0.501	0.501	F=2.278
Officer Hispanic	302	0.278	0.449	397	0.237	0.426	F=1.546
Officer Male	302	0.927	0.26	397	0.95	0.219	F=1.532
Officer Age	302	37.169	7.555	397	36.693	6.957	F=0.746
Experience (Yrs)	302	9.841	6.264	397	9.637	5.818	F=0.197
Prior Incidents	302	7.45	8.703	397	8.073	9.521	F=0.79
In Uniform	302	0.401	0.491	397	0.499	0.501	F=6.695***
Rank: Officer	302	0.897	0.304	397	0.884	0.32	F=0.305
Officer Injured	302	0.126	0.332	397	0.212	0.409	F=8.842***
Officer Fired Handgun	302	0.947	0.224	397	0.982	0.132	F=6.781***
Officer Fired TASER	302	0	0	397	0	0	
Number of weapons discharged	302	1.01	0.129	397	1.02	0.141	F=0.974
Subject Characteristics							
Subject Black	302	0.752	0.433	397	0.786	0.411	F=1.138
Subject White	302	0.026	0.161	397	0.05	0.219	F=2.547
Subject Hispanic	302	0.182	0.387	397	0.149	0.356	F=1.409
Subject Male	302	0.974	0.161	397	0.987	0.112	F=1.814
Subject Age	302	27.162	10.126	397	27.889	9.763	F=0.921
Subject Armed	302	0.685	0.465	397	0.884	0.32	F=44.619***
Domestic charge	302	0.026	0.161	397	0.06	0.239	F=4.546**
Drug charge	302	0.225	0.418	397	0.141	0.349	F=8.391***
Assault charge	302	0.679	0.468	397	0.69	0.463	F=0.103
Weapon charge	302	0.358	0.48	397	0.353	0.478	F=0.018
Weapon: None	302	0.298	0.458	397	0.123	0.329	F=34.335***
Weapon: Fists	302	0.046	0.211	397	0.045	0.208	F=0.004
Weapon: Handgun	302	0.397	0.49	397	0.526	0.5	F=11.632***
Weapon: Long Gun	302	0.013	0.115	397	0.023	0.149	F=0.833
Incident Characteristics							
Active resistance	302	1.023	0.973	397	1.073	0.9	F=0.491
Passive resistance	302	2.368	1.345	397	2.728	0.98	F=16.791***
Deadly force (S)	302	0.94	0.237	397	0.992	0.087	F=16.27***
Officer fired first	302	0.788	0.409	397	0.768	0.422	F=0.388
Daylight	302	0.391	0.489	397	0.272	0.446	F=11.189***
Artificial light	302	0.467	0.5	397	0.627	0.484	F=18.29***
Darkness	302	0.003	0.058	397	0	0	F=1.315
Dawn	302	0.007	0.081	397	0.013	0.112	F=0.616
Clear weather	302	0.887	0.317	397	0.899	0.301	F=0.253
Cloudy weather	302	0.003	0.058	397	0	0	F=1.315
Fog/Haze/Smoke weather	302	0	0	397	0.003	0.05	F=0.76
Rain/Snow/Hail weather	302	0.066	0.249	397	0.068	0.252	F=0.009
Indoor	302	0.139	0.347	397	0.128	0.335	F=0.167

Note: Time period is 2004–2019. "Prior Incidents" is a variable that equals the count of previous tactical response report incidents for a given officer at the time of the shooting. "Passive resistance" are counts of passive resistance in an incident (e.g. failure to comply with a verbal command). " Active resistance" are counts of active resistance in an incident (e.g. physically resisting arrest). "Deadly force (S)" is a binary variable that takes 1 if a subject used deadly force against an officer. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A2: Effect of OIS involving firearms on crime reporting adjusting for small number of clusters

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured x OIS	-0.037**	-0.036**	-0.037**	-0.031***	-0.028***
	(0.015)	(0.015)	(0.014)	(0.010)	(0.010)
	[-0.069, -0.005]	[-0.069, -0.004]	[-0.068, -0.004]	[-0.054, -0.008]	[-0.049, -0.007]
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. Fast wild bootstrap confidence intervals with Rademacher weights and 999 iterations presented in square brackets. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A3: Effect of OIS involving firearms and TASERs on crime reporting

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.006	0.007	0.008	-0.003	-0.000
	(0.007)	(0.007)	(0.007)	(0.005)	(0.004)
OIS	0.018***	0.019***	0.019***	0.009**	0.007**
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Subject Injured x OIS	-0.004	-0.005	-0.006	-0.003	-0.004
	(0.008)	(0.008)	(0.007)	(0.006)	(0.004)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	60,365	60,365	60,365	60,365	60,365
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A4: Effect of OIS involving firearms on crime reporting (oldest officer, oldest subject)

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.012	0.009	0.009	0.004	0.009
	(0.013)	(0.012)	(0.011)	(0.010)	(0.010)
OIS	0.027**	0.025**	0.023**	0.021***	0.019**
	(0.011)	(0.011)	(0.009)	(0.007)	(0.008)
Subject Injured x OIS	-0.032**	-0.032**	-0.033**	-0.030***	-0.027**
	(0.015)	(0.015)	(0.013)	(0.010)	(0.010)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A5: Effect of OIS involving firearms on crime reporting (youngest officer, oldest subject)

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.020	0.015	0.013	0.006	0.009
	(0.014)	(0.014)	(0.013)	(0.010)	(0.010)
OIS	0.032***	0.031***	0.029**	0.025***	0.021**
	(0.011)	(0.011)	(0.010)	(0.008)	(0.009)
Subject Injured x OIS	-0.038**	-0.039**	-0.038**	-0.031***	-0.027**
	(0.016)	(0.016)	(0.015)	(0.011)	(0.011)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,111	10,111	10,111	10,111	10,111
Clusters	22	22	22	22	22
Mean (dep. var)	47.212	47.212	47.212	47.212	47.212

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A6: Effect of OIS involving firearms on crime reporting (youngest officer, youngest subject)

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.020	0.016	0.013	0.007	0.009
	(0.013)	(0.013)	(0.011)	(0.009)	(0.008)
OIS	0.030***	0.029***	0.027***	0.023***	0.021**
	(0.010)	(0.010)	(0.009)	(0.007)	(0.008)
Subject Injured x OIS	-0.035**	-0.035**	-0.035**	-0.029***	-0.028***
	(0.015)	(0.015)	(0.014)	(0.010)	(0.009)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,111	10,111	10,111	10,111	10,111
Clusters	22	22	22	22	22
Mean (dep. var)	47.212	47.212	47.212	47.212	47.212

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A7: Testing sensitivity of estimates to pre and post periods

	ln(Reported Crime + 0.1)				
Pre:	21 days	14 days	10 days	14 days	21 days
Post:	14 days	10 days	14 days	21 days	21 days
	(1)	(2)	(3)	(4)	(5)
Subject Injured x OIS	-0.026***	-0.034***	-0.037***	-0.029***	-0.019**
	(0.008)	(0.010)	(0.011)	(0.010)	(0.009)
Observations	10,814	8,469	9,224	11,812	12,763
Clusters	22	22	22	22	22
Mean (dep. var)	46.795	46.795	46.795	46.795	46.795

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include officer, subject, and incident controls as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A8: Effet of OIS involving firearms on crime reporting accounting for ShotSpotter incidents

	ln(Reported Crime)		
	(1)	(2)	(3)
Subject Injured	0.013	0.059	0.002
	(0.012)	(3,192.608)	(3,831.695)
OIS	0.024**	0.013	0.010
	(0.009)	(0.032)	(0.033)
Subject Injured x OIS	-0.031***	0.001	0.002
	(0.010)	(0.091)	(0.089)
ShotSpotter Incidents			0.013**
			(0.004)
Observations	8,344	413	413
Clusters	13	13	13
Mean (dep. var)	51.648	51.648	51.648

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A9: Heterogeneous effect of OIS involving firearms and TASERs on reported crime

	ln(Reported Crime + 0.1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subject Injured x OIS	-0.004	-0.007	-0.008	-0.007	-0.012**	-0.003
	(0.004)	(0.011)	(0.006)	(0.004)	(0.005)	(0.004)
Subject Injured x OIS x Subject Unarmed		0.005				
		(0.013)				
Subject Injured x OIS x Black Subject			0.005			
			(0.008)			
Subject Injured x OIS x Black District				0.008		
				(0.008)		
Subject Injured x OIS x Different Races					0.013**	
					(0.006)	
Subject Injured x OIS x Subject Dead						0.007
						(0.057)
Observations	60,365	60,365	60,365	60,365	60,365	60,366
Clusters	22	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(ReportedCrime). All columns include officer, subject, and incident controls as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A10: Effect of OIS involving firearms and armed subjects on reported crime

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.023*	0.018	0.006	-0.002	0.001
	(0.013)	(0.013)	(0.015)	(0.012)	(0.012)
OIS	0.030***	0.029***	0.029**	0.020***	0.021***
	(0.010)	(0.009)	(0.010)	(0.006)	(0.007)
Subject Injured x OIS	-0.037**	-0.036**	-0.032**	-0.021*	-0.021*
	(0.015)	(0.015)	(0.015)	(0.012)	(0.011)
Subject Unarmed		-0.036**	-0.056**	-0.047***	-0.013
		(0.017)	(0.026)	(0.012)	(0.014)
Subject Injured x Subject Unarmed			0.073*	0.070***	0.072***
			(0.038)	(0.019)	(0.023)
OIS × Subject Unarmed			-0.009	0.009	-0.003
			(0.018)	(0.013)	(0.012)
Subject Injured x OIS x Subject Unarmed			-0.032	-0.057**	-0.048**
			(0.028)	(0.022)	(0.020)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A11: Effect of OIS involving firearms and Black subjects on reported crime

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.023*	0.018	0.048	0.008	0.014
	(0.013)	(0.013)	(0.030)	(0.017)	(0.020)
OIS	0.030***	0.029***	0.046***	0.034***	0.038***
	(0.010)	(0.009)	(0.014)	(0.009)	(0.010)
Subject Injured x OIS	-0.037**	-0.036**	-0.067**	-0.049**	-0.047*
	(0.015)	(0.015)	(0.024)	(0.021)	(0.023)
Subject Unarmed		-0.036**	-0.041**	-0.030***	-0.000
		(0.017)	(0.018)	(0.009)	(0.009)
Subject Injured x Black Subject			-0.041	0.003	-0.002
			(0.033)	(0.017)	(0.025)
Subject Injured x OIS x Black Subject			0.043	0.025	0.027
			(0.030)	(0.023)	(0.025)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A12: Effect of OIS involving firearms in majority Black districts on reported crime

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.043*	0.038*	0.039*	0.013	0.028*
	(0.021)	(0.020)	(0.020)	(0.017)	(0.016)
OIS	0.049***	0.047***	0.042**	0.026*	0.030**
	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)
Subject Injured x OIS	-0.064**	-0.062**	-0.058**	-0.039*	-0.038**
	(0.024)	(0.023)	(0.022)	(0.020)	(0.018)
OIS × Black District	-0.031	-0.029	-0.024	-0.005	-0.016
	(0.020)	(0.018)	(0.018)	(0.015)	(0.015)
Subject Injured x Black District	-0.034	-0.033	-0.034	-0.004	-0.021
	(0.025)	(0.021)	(0.021)	(0.016)	(0.019)
Subject Injured x OIS x Black District	0.046	0.042	0.035	0.012	0.017
	(0.028)	(0.027)	(0.026)	(0.021)	(0.021)
Black District				-0.009	
				(32,857.696)	
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A13: Effect of OIS involving firearms and White-Black officer-subjects on reported crime

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.023*	0.016	0.011	0.004	0.011
	(0.013)	(0.013)	(0.024)	(0.016)	(0.012)
OIS	0.030***	0.029***	0.030**	0.030***	0.025**
	(0.010)	(0.009)	(0.012)	(0.009)	(0.010)
Subject Injured x OIS	-0.037**	-0.035**	-0.044**	-0.039**	-0.034*
	(0.015)	(0.015)	(0.020)	(0.017)	(0.018)
Subject Injured × diff_raceYes			0.009	0.012	0.002
			(0.035)	(0.017)	(0.017)
Subject Injured x OIS x Different Races			0.016	0.011	0.014
			(0.028)	(0.019)	(0.022)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A14: Effect of OIS involving firearms and fatally injured subjects on reported crime

	ln(Reported Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.023*	0.018	0.013	0.009	0.013
	(0.013)	(0.013)	(0.013)	(0.011)	(0.012)
OIS	0.030***	0.029***	0.027***	0.024***	0.020**
	(0.010)	(0.009)	(0.008)	(0.006)	(0.007)
Subject Injured x OIS	-0.037**	-0.037**	-0.034**	-0.028**	-0.025**
	(0.015)	(0.015)	(0.016)	(0.012)	(0.010)
Subject Dead			-0.020	0.037	-0.002
			(0.033)	(0.053)	(0.069)
Subject Injured x Subject Dead			0.065	-0.022	0.019
			(0.044)	(0.056)	(0.071)
OIS × Subject Dead			-0.015	-0.106**	-0.087
			(0.049)	(0.050)	(0.056)
Subject Injured x OIS x Subject Dead			-0.020	0.080	0.060
			(0.057)	(0.055)	(0.058)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,277	10,277	10,277	10,277	10,277
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A15: Effect of OIS involving firearms on reported serious crime

	ln(Reported Serious Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.024	0.018	0.017	0.007	0.002
	(0.018)	(0.020)	(0.019)	(0.017)	(0.014)
OIS	0.029*	0.027*	0.024*	0.013	0.013
	(0.014)	(0.014)	(0.012)	(0.011)	(0.011)
Subject Injured x OIS	-0.041*	-0.041*	-0.042*	-0.029	-0.022
	(0.022)	(0.023)	(0.023)	(0.018)	(0.017)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A16: Effect of OIS involving firearms on reported violent crime

	ln(Reported Violent Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	-0.011	-0.013	-0.018	-0.029	0.017
	(0.031)	(0.032)	(0.036)	(0.032)	(0.040)
OIS	0.065**	0.060*	0.052	0.038	0.046
	(0.030)	(0.031)	(0.031)	(0.030)	(0.034)
Subject Injured x OIS	-0.043	-0.042	-0.038	-0.018	-0.029
	(0.045)	(0.047)	(0.048)	(0.042)	(0.048)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A17: Effect of OIS involving firearms on reported less serious crime

	ln(Reported Less Serious Crime + 0.1)				
	(1)	(2)	(3)	(4)	(5)
Subject Injured	0.023	0.021	0.020	0.013	0.026*
	(0.014)	(0.013)	(0.012)	(0.010)	(0.015)
OIS	0.032**	0.033***	0.030**	0.030***	0.025**
	(0.012)	(0.011)	(0.011)	(0.008)	(0.011)
Subject Injured x OIS	-0.036**	-0.036**	-0.035**	-0.034***	-0.034***
	(0.016)	(0.016)	(0.014)	(0.010)	(0.011)
Officer Controls	✓	✓	✓	✓	✓
Subject Controls		✓	✓	✓	✓
Incident Controls			✓	✓	✓
Month FEs				✓	✓
District × Year FEs					✓
Observations	10,276	10,276	10,276	10,276	10,276
Clusters	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Crime + 0.1). All columns include district and year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A18: Heterogeneous effects of OIS involving firearms on reported serious crime

	ln(Reported Serious Crime + 0.1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subject Injured x OIS	-0.022	-0.021	-0.020	-0.025	-0.027	-0.020
	(0.017)	(0.015)	(0.018)	(0.031)	(0.027)	(0.025)
Subject Injured x OIS x Subject Unarmed		-0.007				
		(0.032)				
Subject Injured x OIS x Subject Dead			0.095**			
			(0.045)			
Subject Injured x OIS x Black Subject				0.004		
				(0.037)		
Subject Injured x OIS x Black District					0.008	
					(0.033)	
Subject Injured x OIS x Different Races						-0.000
						(0.034)
Observations	10,276	10,276	10,277	10,276	10,276	10,276
Clusters	22	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Serious Crime + 0.1). Serious crimes are UCR Part I Index crimes. All columns include officer, subject, and incident controls as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A19: Heterogeneous effects of OIS involving firearms on reported violent crime

	ln(Reported Violent Crime + 0.1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subject Injured x OIS	-0.029	-0.034	-0.012	-0.065	-0.058	-0.123*
	(0.048)	(0.052)	(0.047)	(0.079)	(0.062)	(0.071)
Subject Injured x OIS x Subject Unarmed		-0.025				
		(0.086)				
Subject Injured x OIS x Subject Dead			-1.310***			
			(0.200)			
Subject Injured x OIS x Black Subject				0.052		
				(0.094)		
Subject Injured x OIS x Black District					0.048	
					(0.082)	
Subject Injured x OIS x Different Races						0.189*
						(0.105)
Observations	10,276	10,276	10,277	10,276	10,276	10,276
Clusters	22	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Violent Crime + 0.1). Violent crimes are the following UCR Part I index crimes–homicide, rape, aggravated assault, and robbery. All columns include officer, subject, and incident controls as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE A20: Heterogeneous effects of OIS involving firearms on reported less serious crime

	ln(Reported Less Serious Crime + 0.1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subject Injured x OIS	-0.034***	-0.024*	-0.032**	-0.063*	-0.049**	-0.046**
	(0.011)	(0.012)	(0.012)	(0.032)	(0.021)	(0.022)
Subject Injured x OIS x Subject Unarmed		-0.065**				
		(0.027)				
Subject Injured x OIS x Subject Dead			0.069			
			(0.074)			
Subject Injured x OIS x Black Subject				0.041		
				(0.038)		
Subject Injured x OIS x Black District					0.026	
					(0.025)	
Subject Injured x OIS x Different Races						0.026
						(0.032)
Observations	10,276	10,276	10,277	10,276	10,276	10,276
Clusters	22	22	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193	47.193	47.193

Note: Dependent variable is ln(Reported Less Serious Crime + 0.1). Less serious crimes are UCR Part II index crimes. All columns include officer, subject, and incident controls as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A21: Effect of OIS involving firearms and TASERs on types of reported crime

	All	Serious	Violent	Less Serious
	(1)	(2)	(3)	(4)
Subject Injured	-0.000	-0.001	0.010	-0.001
	(0.004)	(0.007)	(0.016)	(0.004)
OIS	0.007**	0.005	0.025**	0.007*
	(0.003)	(0.004)	(0.010)	(0.004)
Subject Injured x OIS	-0.004	-0.004	-0.018	-0.002
	(0.004)	(0.007)	(0.019)	(0.005)
Observations	60,365	60,365	60,365	60,365
Clusters	22	22	22	22
Mean (dep. var)	47.193	47.193	47.193	47.193

Note: Dependent variable in column 1 is ln(Reported Crime + 0.1), in column 2 is ln(Reported Serious Crime + 0.1), in column 3 is ln(Reported Violent Crime + 0.1), and in column 4 is ln(Reported Less Serious Crime + 0.1). Serious crimes are UCR Part I index crimes. Violent crimes are a subset of UCR Part I index crimes: homicide, rape, robbery, and aggravated assault. Less serious crimes are UCR Part II index crimes. All columns include district and year fixed effects. All columns include officer, subject, and incident controls as well as district, year, month, and district-year fixed effects. Standard errors clustered at the district level. *** p < 0.01, ** p < 0.05, * p < 0.1

Chapter 3

On the Effects of COVID-19 Safer-At-Home Policies on Social Distancing, Car Crashes and Pollution

3.0.1 Abstract

This paper investigates the impacts of COVID-19 safer-at-home polices on collisions and pollution. We find that statewide safer-at-home policies lead to a 20% reduction in vehicular collisions and that the effect is entirely driven by less severe collisions. For pollution, we find particulate matter concentration levels approximately $1.5\mu g/m^3$ lower during the period of a safer-at-home order, representing a 25% reduction. We document a similar reduction in air pollution following the implementation of similar policies in Europe. We calculate that as of the end of June 2020, the benefits from avoided car collisions in the U.S. were approximately $16 billion while the benefits from reduced air pollution could be as high as $13 billion.

3.0.2 Thanks

We thank Mohammad Elfeitori, Ramanvir Grewal and Kelly Liu for their excellent research assistance. We declare no conflict of interest and no research funding.

3.1 Introduction

The emergence of COVID-19 (formally termed SARS-CoV-2 by the International Committee on Taxonomy of Viruses) has fundamentally changed human behavior. Characterized as a pandemic by the World Health Organization on March 11, 2020, the global scientific community is actively researching the virus and its impact. As of December 31, 2020, the United States has seen over 341 thousand deaths and over 19 million confirmed cases. COVID-19 has led most state governors to impose safer-at-home orders. We consider safer-at-home orders to be a blanket term that captures other efforts introduced simultaneously (or nearly so) whose goal was the suspension of economic activity and public interaction in an effort to "flatten the curve".[1] To date, the focus of the debate over safer-at-home orders has been on their efficacy of transmission suppression and the implicit trade-off between lives saved and reduced economic activity. Safer-at-home orders have resulted in serious negative

[1]See Hsiang et al., 2020 and Sanga and McCrary, 2020 for similar treatment of safer-at-home orders and lockdowns.

impacts on Americans in several dimensions: studies have shown large impacts on the labor market, mental health, and domestic violence incidents, for instance (e.g., Adams-Prassl et al., 2020; Beland, Brodeur, and Wright, 2020; Brodeur et al., 2021; Leslie and Wilson, 2020).

In this paper, we study other potential *positive* externalities of safer-at-home orders; a decrease in air pollution and automobile collisions.[2] We rely on a difference-in-differences framework for identification. The setting is attractive for at least two reasons. First, not all states (nor all counties) implemented safer-at-home orders, and there is significant variation in implementation timing for those who do. Second, our identification strategy allows us to address issues of reverse causality and omitted variables bias by comparing states (or counties) that implemented safer-at-home orders at different points in time. Our identification assumption is that, conditional on COVID-19 incidence and other policies implemented (e.g., statewide face masks mandates), the difference in pollution (or automobile collisions) between areas with and without safer-at-home orders would be constant over time.

We first investigate the impact of safer-at-home orders on pollution. We find that state safer-at-home policies decreased air pollution (specifically PM2.5) by almost 25%, with larger effects for urban counties. This large effect size suggests that these policies reduce emissions by almost one half of a within-county standard deviation. Our estimates also suggest the issuance of a state order reduces the number of county-days with an acceptable PM2.5 level by around 10 percentage points (nearly eliminating 'polluted' days). Further, we find that the decrease in air pollution persists for weeks after the order is lifted.

We check whether our results hold in a different setting by studying the impact of countrywide 'lockdowns' policies in Europe. More specifically, we build a data set of pseudo 'counties' for France, Germany, Italy, Spain, and the United Kingdom, and document a (temporary) decrease in air pollution following the implementation of lockdowns.

In addition, we estimate if there are car collision externalities stemming from the safer-at-home orders. We use daily-level traffic collision data for all counties, originally sourced through two real-time maps service API's which draw from sources such as law enforcement and transportation departments. We identify a large reduction - about 20% - in traffic collisions after a state order is issued. We also examine their severity using a four-point scale index based on traffic flow/disruption. Following an order, we find a large and significant decrease in the most common types of collisions, with an increase in the minority of the most serious. Using social mobility data, we provide evidence that individuals shift their travel away from traditionally congested periods which helps explain the pattern of overall reductions but increases in collision severity.

We then explore the heterogeneous effects of safer-at-home policies on pollution and collisions across county characteristics. We find that the decline in pollution and collisions from safer-at-home orders is larger in urban counties. Our results also indicate that counties in states with a larger share of occupations that can be done remotely experience a larger reduction in pollution and collisions from these policies.

[2]Researchers have also begun investigating the direct effect of air pollution on COVID-19. For example, higher levels of contemporaneous air pollution were associated with increased transmission and mortality in the Chinese context by Zhang, Xue, and Jin, 2020. This correlation is also investigated for Italy, Spain, France, Germany, U.K. and U.S. in Pansini and Fornacca, 2020. If this result were to hold in our context, the benefits of safer-at-home orders reducing air pollution may also include direct reductions in COVID-19 mortality and transmission.

Lastly, we provide some back-of-the-envelope calculations of the estimated benefits from reduced pollution and collisions from safer-at-home orders. Using previous estimates of willingness to pay for pollution reduction from the U.S. we find the benefit from reduced pollution ranges from \$154 million to \$13 billion as of the end of June 2020. Using estimates of the societal cost of car collisions from the National Highway Traffic Safety Administration gives approximately \$16 billion in costs avoided as a result of safer-at-home orders as of June 2020.

We contribute to a growing literature informing the ongoing debate about safer-at-home orders (see Brodeur et al., 2020 for a literature review). Previous studies have documented the positive impact of lockdowns on (preventing) COVID-19 incidence, but also their potential negative effects on the economy (e.g., Kong and Prinz, 2020), domestic violence (e.g., Leslie and Wilson, 2020), child maltreatment (e.g., Bullinger et al., 2020) and mental health (e.g., Adams-Prassl et al., 2020) among other socioeconomic dimensions. We contribute to this literature by pointing out two unintended benefits of lockdowns: decreased car crashes and reduced air pollution.

The most relevant papers to ours are possibly He, Pan, and Tanaka, 2020 and Dang and Trinh, Forthcoming. He, Pan, and Tanaka, 2020 provide evidence that lockdowns in China decreased PM2.5 by approximately 25%. Dang and Trinh, Forthcoming provide cross-national evidence for 164 countries on the impact of COVID-19 lockdowns on global concentration of NO_2 and PM2.5. They find that lockdowns decreased NO_2 and PM2.5 levels by about 5 percent. Other relevant work includes Cicala et al., 2020, Graf, Quaglia, and Wolak, Forthcoming and Le Quéré et al., 2020. Cicala et al., 2020 provide evidence that electricity consumption fell in the U.S. during the pandemic.[3] Graf, Quaglia, and Wolak, Forthcoming investigate how lockdowns can affect electricity market performance using Italian data. Le Quéré et al., 2020 study the effects of government policies on country-level energy demand finding CO_2 emissions (estimated directly from confinement data) decreased by 17% compared to previous year levels. Our paper focuses on the immediate positive externalities from a reduction in road congestion and ambient particulate matter, using the available local and real time data necessary to understand how local and state government policies affect behavior.

The rest of the paper is organized as follows. Section 3.2 details the data, while Section 3.3 describes our identification strategy. We discuss the impacts of safer-at-home policies on pollution in Section 3.4. Section 3.5 investigates the effects of safer-at-home policies on collisions. In Section 3.6, we investigate the relationship between mobility and collision during safer-at-home orders. Section 3.7 provides our back of the envelope calculations of the positive externalities these policies. Section 3.8 concludes.

3.2 Data

In this section, we describe our data. We first provide information on COVID-19 cases and fatalities, and how they vary over time and across states. We then describe data sources for safer-at-home orders and other policies. Last, we describe our pollution and collision data.

[3] Leach, Rivers, and Shaffer, 2020 document a reduction in the closely related Canadian context.

3.2.1 COVID-19 Known Cases and Deaths

The first COVID-19 case in the U.S. was a man who had returned from Wuhan, China to Washington State. The case was confirmed on January 20, 2020. Six additional states confirmed cases later in January and February. The first case of community transmission was confirmed in California, on February 26, 2020. As of April 30, 2020 there were over 1 million confirmed cases due to COVID-19 in the United States. On the last day of 2020, the CDC reported 19,663,976 total cases and 341,199 total deaths from COVID-19.

The COVID-19 known cases and deaths data comes from the Github repository associated with the Johns Hopkins University interactive dashboard. The data are available here: https://github.com/CSSEGISandData/COVID-19. Appendix Figures A1 and A2 illustrate the geographic distribution of COVID-19 known cases and deaths per 10,000 inhabitants, respectively.

3.2.2 Safer-at-Home Policy

Data for safer-at-home policies are from the *New York Times*.[4] Figures 1 and 2 present maps indicating counties and states that implemented a safer-at-home policy prior to April 30, 2020, respectively. Nearly all states had implemented such a policy at this point in time, and the timing of implementation varies considerably. The first state to implement a safer-at-home policy was California on March 19th, 2020. 18 more states followed California in the following week. In 2020, 43 states (including the District of Columbia) had implemented some form of lockdown, representing 2,628 counties. Only California, North Carolina and Ohio had an active safer-at-home order on the last day of 2020. 148 counties implemented a county-level safer-at-home policy, of which 141 are located in states that would eventually have a statewide policy. The median county implemented its safer-at-home policy one week prior to the statewide policy.[5]

We also use data on the stringency of safer-at-home orders from the Oxford COVID-19 Government Response Tracker (OxCGRT) implemented by the University of Oxford's Blavatnik School of Government. We rely on an ordinal scale measure of safer-at-home requirements, which takes the value of zero in the absence of an order, one if governmental authorities recommended a state to stay home, and two if staying-at-home was required with few exceptions such as daily exercise, grocery shopping, and essential trips.

3.2.3 Other Policies

We also gather data on the following COVID-19 statewide policies: day care closures, freezes on eviction, mandatory face mask policies, and mandated quarantine for individuals arriving from another state. Data on the implementation and duration of these policies come from Raifman et al., 2020. We provide more detailed information about these policies in the Appendix 3.11.1.

[4] Data are available at https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html.

[5] Only seven counties (Brazos, Comal, Humboldt, Kings, Mendocino, Merced and Milam) implemented an additional safer-at-home policy after a statewide policy.

3.2.4 Social Distancing Data

We extract data on social distancing cell phone data from Unacast's COVID-19 Toolkit. Unacast provides a Social Distancing Scoreboard at the county-level using cell phone data which aims to empower organizations to evaluate the effectiveness of social distancing initiatives (Brodeur, Grigoryeva, and Kattan, 2020).

Using data pre-COVID-19 outbreak as a baseline, Unacast computes rate of changes in average distance travelled, non-essential visitation, and human encounters. For our analysis, we rely on the first index. Unacast's data is available starting February 24th, 2020 and unavailable for many counties.

3.2.5 Air Pollution and Weather

Particulate Matter Concentrations

Air pollution data is from *in situ* monitors and provided by AirNow, a partnership between United States agencies.[6] The primary pollutant we use in our analysis is particulate matter with diameters less than 2.5 micrometers - particles small enough that they are capable of being inhaled and passing through the blood brain barrier. Emissions from combustion of gasoline, oil, diesel fuel or wood produce much of the PM2.5 pollution found in outdoor air. PM2.5 is associated with the greatest proportion of adverse health effects related to air pollution in the United States.

The Environmental Protection Agency monitors PM2.5 levels to protect public health and the environment, and has found average decreasing trends in the last two decades. From 1990 to 2016, PM2.5 measures had fallen around 25%, a reduction roughly equal to our estimated effect of the safer-at-home orders.

To aggregate PM2.5 levels from the monitor level to the county-level, we assign each county's population weighted centroid to the three nearest air quality monitoring stations. Readings from each station are then averaged using inverse distance weights where closer monitors carry proportionally more weight in the pollution level. We also restrict our estimations to areas where the nearest station is within 50 kilometers (as accurate PM2.5 levels are necessarily local). In our sample, the minimum distance to a pollution monitoring station is 114 meters, while the mean and median are approximately 25km away.

In Appendix Figure A3, we present average weekly PM2.5 levels by county for the week of March 1-7, 2020. This period, prior to any safer-at-home policies serves as a visual representation of the geographic distribution of air pollution levels. In Appendix Figure A4, we present the same PM2.5 measures during the final week of April - after all eventually treated states had implemented a safer-at-home policy. See also Appendix Figure A5 for the distribution of PM2.5 concentrations for all county-days in our sample.

Other Air Pollution Measures

While a derivative of PM2.5 levels, we also provide results using the Air Quality Index (AQI). This unit-less measure ranges from 0 to 500, with a score below 50 representing 'not harmful' levels of air pollution.

We also use aerosol optical depth (AOD) as an alternative and more geographically dispersed measure of air pollution. AOD estimates the amount of aerosol (tiny

[6] The U.S. Environmental Protection Agency, National Oceanic and Atmospheric Administration (NOAA), National Park Service, NASA, Centers for Disease Control, and tribal, state, and local air quality agencies. The centralized system provides uniform quality control and reporting consistency.

solid and liquid particles released by cars, industries, fires, etc.) present in the atmosphere and has been used as a proxy for surface air pollution (such as PM2.5). Technically, AOD measures the "extinction of a ray of light as it passes through the atmosphere" where extinction refers to diminishment either from absorption or scattering. A greater measure of AOD indicates a higher estimate of surface air pollution. Specifically, we use daily estimates of AOD at the 10km × 10km resolution derived from the MODIS platform. We use pre-processed quality-controlled estimates which account for both low-quality measurements and highly reflective surfaces such as deserts.[7] Appendix Figure A6 illustrates the distribution of AOD for all county-days in our sample.

Temperature and Precipitation

For temperature and precipitation, we use the NOAA CPC Global Daily Temperature data set.[8] This data set is typically used for verification of other temperature and precipitation products, and is available from 1979 to the present. This data set provides global coverage of temperature and precipitation at the 0.5 degree × 0.5 degree spatial resolution. At the geographic center of the contiguous United States (39.83 North and 98.58 West), this represents a 55 km × 42 km grid. This data set is typically available in real time. The underlying meteorological data comes from the Global Telecommunication System daily reports. They are from 6,000 to 7,000 global stations, with 10% of those in the United States. The station data is then gridded using the Shepard algorithm.

3.2.6 Collision Data

For our analysis on car crashes, we rely on collision data at the county-level from 49 states from January 1, 2020 to June 30, 2020 created by Moosavi et al., 2019. The data set is built from continuously streaming traffic data from MapQuest and Microsoft Bing map services and includes location, date, and severity of each crash. These services stream traffic incidents captured from national and state departments of transportation, law enforcement, traffic cameras, and traffic sensors. The authors of the data set collected data at 90 second intervals from 6am to 11pm and 150 second intervals from 11pm to 6am. Our main variable of interest is the daily number of collisions per county. The severity of an accident is coded as a number ranging from 1 to 4 with 1 being the smallest impact on traffic and 4 being a significant impact on traffic. Table 1 provides summary statistics for collisions. In our sample, there are about 1.8 collisions per day per county.

3.3 Identification Strategy

Our hypothesis is that safer-at-home policies decreased PM2.5 concentrations and collisions. To investigate this hypothesis, we estimate the following difference-in-differences specification:

[7] At the time of writing, the most recent data available for the AOD data was November 30, 2020, corresponding to day 335 of the year. AOD is not produced where clouds are definitively present. A quick guide to Aerosol Optical Depth can also be found from NOAA-NASA.

[8] NOAA stands for the National Oceanic and Atmospheric Administration, an American scientific agency within the United States Department of Commerce that focuses on the conditions of the oceans, major waterways, and the atmosphere. They manage the United States operational environmental satellites.

$$y_{cst} = \alpha + \beta StateSafer_{st} + \lambda CountySafer_{cst} + \gamma_c + \delta_t + X'_{cst}\omega + \varepsilon_{cst} \quad (1)$$

where y_{cst} is, for instance, daily average PM2.5 measured in $\mu g/m^3$ in county c in state s and year t. We include a full set of county dummies γ_c to control for time-invariant county characteristics and calendar date dummies δ_t (e.g., a separate dummy for March 1, 2020, March 2, 2020, etc.). The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. We thus have an unbalanced panel of counties. For the bulk of our analysis, the sample is restricted to counties that eventually implement a county order or are under a statewide order. The last day in our sample is June 30th, 2020, although the sample ends prior to June 30th for the vast majority of counties. The state-level variable $StateSafer_{st}$ equals one once the state has implemented the order and zero for the pre-policy period.[9] Our primary coefficient of interest is β. We also investigate the impact of county orders. The county-level variable $CountySafer_{ct}$ equals one once the county has issued the order and zero for the pre-policy period. Another coefficient of interest when included is thus λ. We cluster standard errors at the state-level, corresponding to the primary policy and treatment level.

Note that the adoption of safer-at-home policies and timing of adoption may be endogenously related to the severity of the virus. We thus include, X_{cst}, a vector of county-day level covariates including known COVID-19 cases and deaths per 10,000 inhabitants. County population data comes from 2019 Census estimates. We also control for county-day precipitation and average temperature. Further we include controls for the following statewide policies: day care closures, freezes on eviction, mandatory face mask policies, and mandated quarantine for individuals arriving from another state. The inclusion of these other policy variables help us to identify the impact of safer-at-home policies rather than the joint impact of multiple government interventions.

Our identification assumption is that, conditional on the included control variables, the evolution of PM2.5 concentrations or collisions for counties with safer-at-home policies would not have been different from those without the policies. This amounts to an assumption of parallel trends in PM2.5 or collisions for treated and untreated counties.[10]

Recent research on two-way fixed-effects (TWFE) estimators, which are usually motivated as difference-in-differences with multiple time periods, has identified issues that arise in the presence of heterogeneous treatment effects across groups or time (Callaway and Sant'Anna, Forthcoming; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon and Marcus, 2020). Using the `twowayfeweights` Stata package detailed in De Chaisemartin and d'Haultfoeuille, 2020, we document that 44% of the naive average treatments on the treated for pollution are assigned negative weights, indicating the need to reexamine our results using an alternative estimator. We use the alternative estimators provided by Callaway and Sant'Anna, Forthcoming; Callaway and Sant'Anna, 2020 that are most appropriate for the staggered adoption of safer-at-home policies in our sample. We discuss these alternative estimates

[9] Using the date of announcement instead of the date of implementation yields similar conclusions. State orders are announced an average of three days prior to implementation.

[10] In a further analysis presented in the Appendix, we check the robustness of our results using simulations constructed by applying synthetic control methods to match counties based on pre-policy pollution levels. Our results are quantitatively similar, and the corresponding placebos-in-place behave as expected.

in section 3.4.[11]

3.4 Safer-at-Home Orders and Pollution

In this section, we present the main results for air pollution using our difference-in-differences strategy. We then then provide additional results for Europe and heterogeneity analyses.

3.4.1 Main Results

In Table 2, we present our main result: a state's implementation of a safer-at-home order significantly lowers air pollution in its constituent counties.

This table presents estimates of Equation 1 in which we compare counties in states with and without a statewide safer-at-home policy.[12] In all columns the dependent variable is PM2.5 concentration. We use a total of 1,592 counties in our estimation.

In the first column, we include only date and county fixed effects. The estimated reduction in PM2.5 from a statewide safer-at-home order is statistically significant at the 1% level and suggests that the introduction of a state order reduces PM2.5 levels by 1.7 $\mu g/m^3$. With a mean of the dependent variable of 6.3 $\mu g/m^3$ for the same time period one year prior, this suggests the policy decreased air pollution by more than 25%. For additional context, the within-county standard deviation of PM2.5 during the same period one year prior was 2.8 $\mu g/m^3$, suggesting the policy reduces emissions by more than one *half* of a standard deviation.

In the second column, we include the county's number of confirmed COVID-19 cases per 10,000 inhabitants as a control. In the third column, we include the county's number of confirmed COVID-19 deaths per 10,000 inhabitants as a control. The inclusion of these two controls has almost no effect on the magnitude and significance of our estimates. The estimate for COVID-19 cases is negative and significant (statistically insignificant once we control for COVID deaths), suggesting that increased cases reduces PM2.5 in a county. In contrast, the estimate for COVID-19 deaths are not statistically significant.

In the fourth column, we control for four distinct statewide policies, while we add local weather conditions - temperature and precipitation - in the fifth column. Overall, the inclusion of other policies has little effect on the magnitude and significance of our state order estimates on PM2.5. The inclusion of weather controls makes the reductive effect of safer-at-home orders on air pollution smaller in magnitude, but the estimate remains large and significant. We note that temperature (precipitation) is positively (negatively) related to PM2.5 levels.

In Appendix Table A2, we test whether counties that implemented an order prior to a statewide order are differently affected than counties that did not implement an order. In columns 1 and 2, we reproduce our main result for all counties. Columns 3 and 4 restrict the sample to a subset of counties that issued their own orders prior to their respective state. In columns 5 and 6, we restrict the sample to counties that did not implement an order prior to the statewide order. Overall, we find that the decrease in air pollution following the implementation of a statewide order is large

[11] We do not provide staggered adoption estimates for collisions as these estimators are not presently available for count data, however given that the share of negative weights is similar to pollution we are optimistic that they would behave similarly to those of pollution.

[12] Our main results are robust to conducting the analysis at the state-level rather than at the county-level. See Appendix Table A1 for the estimates.

and significant for the two sets of counties. The estimate of the effect of a county order is negative, but barely statistically insignificant, suggesting that conditional on a state order being present, there is minimal *additional* reduction in PM2.5 from an additional county order being in effect.

In Appendix Table A3, we re-estimate the unweighted estimates from Table 2 and now weight by county population. In this manner, we place more emphasis on relatively more populous (and concurrently polluted) counties. The inclusion of this weighting increases the effect estimate by about 50% - the estimated reduction grows from 1.7 to 2.5 $\mu g/m^3$ - suggesting that the reduction in PM2.5 was larger for more populous counties.

3.4.2 Robustness Checks

In Table 3, we conduct a similar but distinct analysis. The column specifications and structure remains the same as in Table 2. While in Table 2 the dependent variable was PM2.5 concentration, in Table 3 the dependent variable is an indicator that takes a value of one if the PM2.5 level in that county, and on that day, is above the National Ambient Air Quality Standard of 12 $\mu g/m^3$.[13] This reflects a change in interpretation from linear decreases in pollution concentration to decreases in exposure to environmental *hazards*. The interpretations of the coefficients change as well. For example, in column 1 nearly 90% of days prior to a state order have an acceptable level of ambient air pollution. Said differently, around one in ten county-days exceed the tolerable level set by the EPA. When state orders are introduced, almost all days for all counties have acceptably clean air (conditional on our control variables). The incremental addition of controls does not perturb this estimate significantly.[14]

Similarly, we rely on a remotely-sensed measure of air pollution in Appendix Tables A6 and A7; aerosol optical depth (AOD). AOD is a measure of the diffusion of light through the atmospheric column and a popular measure of air pollution which has the distinct advantage of being measured from space. Our main results (for example those presented in Table 3) are restricted to areas within 50 kilometers of an air pollution monitoring site, reflecting the concern that air pollution concentrations are inherently local. The locations of air quality monitoring sites are also most likely non-random to urban centers and, more recently, concerns of their strategic placement have also arisen. A uniformly measured and spatially available measure of air pollution found in aerosol optical depth does not suffer the same weaknesses, but does come with its own. For example, the relationship between AOD and PM2.5 (and other pollutants) is often area-specific and can vary with surface albedo, making aggregation necessary in our setting, difficult. Lower AOD values are typically associated with lower levels of air pollution.

Appendix Table A6 uses aerosol optical depth as the dependent variable, but to maintain a comparison with our main estimates in Table 2, we restrict the sample to counties within 50 km of a monitoring station. In Appendix Table A7, we relax this restriction. The estimates presented in Appendix Tables A6 and A7 suggest that, during a lockdown, the measured within-county AOD is much lower (although

[13] The Environmental Protection Agency sets this standard as "Exposure to fine particle pollution can cause premature death and harmful cardiovascular effects such as heart attacks and strokes, and is linked to a variety of other significant health problems." For example, Bowe et al., 2019 show that among a cohort of U.S. veterans, nine causes of death were associated with PM2.5 exposure above standards set by the Environmental Protection Agency.

[14] In Appendix Tables A4 and A5 we use instead Air Quality Index (AQI) and an indicator for an acceptable Air Quality Index of below 50, respectively. Estimates are substantively the same.

imprecisely estimated), corresponding to reductions in air pollution.[15] The main benefit of AOD over local PM2.5 concentrations is greater spatial availability. But one drawback is the use of particularly coarse readings from the MODIS-TERRA platform and the often relatively clean air in the United States (causing measures of AOD to be often missing). Of course, there is also the possibility that air pollution above the surface level is also being measured.

Appendix Table A8 provides the three aggregated alternative TWFE estimators described in Callaway and Sant'Anna, Forthcoming. Column 1 recreates the estimate from column 1 of Table 2. Column 2 provides the weighted (by group size) average of all estimated county-day average treatment effects. In column 3, we provide the average treatment effect over all lengths of exposure to safer-at-home orders, while column 4 provides the average effect of implementing a safer-at-home order for counties that were under an order in any period. All columns include county and date fixed effects. These alternative estimators confirm that safer-at-home orders have a negative and statistically significant effect on PM2.5.

3.4.3 Graphical Evidence

We provide a visual representation of the reduction in pollution from the implementation of a safer-at-home order in Figure 3.[16] This figure plots estimated PM2.5 levels at daily intervals pre- and post-statewide order. The associated equation is:

$$y_{cst} = \alpha + \sum_{n=-40}^{100} \beta_n (DaysSinceLockdown = n) + \gamma c + \delta_t + X'_{cst}\omega + \varepsilon_{cst} \tag{2}$$

This specification decomposes the level of PM2.5 by the number of days before and since the state order. The regression includes county and date fixed effects in addition to our full set of weather and policy controls. We plot the estimated difference between PM2.5 levels compared to the date that the state order was implemented (which is set to zero). The time window is 40 days before to 100 days after the policy is implemented. The dashed lines represent robust 95% confidence intervals.

Figure 3 shows that PM2.5 levels, within-county and conditional on our covariates, were relatively stable from 40 days before up to the implementation of the statewide order. We begin to see a slight reduction in PM2.5 emissions once the order is implemented, with a much steeper decrease in the weeks following implementation. The negative impact is at its largest at the end of the time window, suggesting that a safer-at-home order's impact on air pollution persists in the weeks following the moment orders are lifted (i.e., all lifted after day *(t+80)*).

3.4.4 Strictness of Order and Heterogeneity Analyzes

In Appendix Table A9, we test whether the estimated air pollution reduction during a safer-at-home order is related to the strictness of the order. Our variables of interest are dummies for "low" safer-at-home order intensity and "high" safer-at-home order intensity. Our estimates show that higher intensity orders are associated with

[15] Of note, once we add precipitation as a control, the point estimate becomes positive for the restricted sample. This is potentially due to the fact that precipitation could affect the estimates of AOD by inadvertently removing days with haze mistaken for cloud cover.

[16] See Appendix Figure A7 for leads and lags of the dependent variable Air Quality Index (AQI).

a larger decrease in air pollution than lower intensity orders. This result has important policy implications since the strictness of orders may be related to many factors such as economic losses and mental health distress.

We investigate whether the magnitude of the documented effect of safer-at-home policies on pollution is related to county-level characteristics in Appendix Table A10. More precisely, we test whether more urban, younger, and Democrat counties experienced a larger decrease in PM2.5 following the implementation of a safer-at-home order.[17] In columns 1 and 2, we split the sample for over and below county urbanization of 50%, respectively. We find that the estimated decrease is about 25% larger for urban counties than rural counties, confirming our previous finding that more populous counties are more affected by state orders.

In columns 3 and 4, we restrict the sample for counties in which a majority of voters voted for President Trump during the 2016 Presidential Election. Column 4 restricts the sample to the other counties. We find that the decrease in pollution is much smaller in counties that supported President Trump in 2016. This finding is in line with Engle, Stromme, and Zhou, 2020, who document that counties with a lower share of votes for Republicans comply more with safer-at-home orders. We confirm this pattern in Appendix Table A11. Of note, we find that 'Trump' counties are also different in many other aspects that may be correlated with the ability to reduce emissions. Readers should therefore be careful when interpreting the findings of this heterogeneity analysis.

Column 5 (6) restricts the sample to counties with relatively more (less) individuals aged at least 65 years old (split by median). We find that counties with relatively more young people experienced a larger decrease in PM2.5, perhaps due to more work being done remotely during-lockdown.

We explore this possibility in columns 7 and 8. We split the sample into counties within states which have above and below median shares of occupations that can be done from home. These classifications of the feasibility of working from home in a given occupation come from Dingel and Neiman, 2020 and associated coding provided by those authors and Ole Agersnap. Column 7 (8) corresponds to counties in states with an above (below) median share of occupations able to be done from home. We find that counties in states with a greater ability to work from home experience a slightly larger decrease in PM2.5.

3.4.5 Europe

We now explore the impact of countrywide lockdowns on pollution in Western Europe. This exercise serves at least two purposes. First, examining the European case is worthy of study in itself. Second, it serves as a test of external validity for our U.S. results. At the time of writing, some form of lockdown had been applied to the residents of most European countries - we focus on France, Germany, Italy, Spain, and the United Kingdom, the largest countries in Europe by population.

For this analysis, we examine how national-level orders affected PM2.5 concentrations at the sub-national level. We divided each country into administrative units at a similar administrative level to United States counties for comparability. A total of 841 areas are used with 96 from France (roughly corresponding to départments), 403 from Germany (roughly corresponding to kreise), 113 from Italy (roughly corresponding to provinces), 55 from Spain (roughly corresponding to provinces) and 192 from the United Kingdom (roughly corresponding to counties).

[17]Data on the share of urban population is based 2010 Census data. Urbanization rate comes from the American Community Survey (ACS-5 years estimates).

Air pollution data was provided by the European Environmental Agency. Pollution remains measured in $\mu g/m^3$ at the population-weighted centroid for each administrative unit (determined in Hall et al., 2019). The mean PM2.5 concentration during the period is $8.7\mu g/m^3$, and the within-unit standard deviation is 6.0. Temperature and precipitation use the same dataset from our U.S. estimates.

Data on lockdowns come from the Coronavirus Government Response Tracker produced by the University of Oxford. This dataset tracks worldwide national (and in rare cases) subnational government responses to the COVID-19 pandemic. We chose this data set because of its uniform coding over the different countries we examine, and a centralized and curated count of confirmed cases and deaths. Further, the uniformity of the mask mandate, quarantine mandate and school closures coding between countries is also valuable. The main drawback to this data set is that the government response, cases, deaths, and policy variables are at the national-level.

Our main results for Europe are presented in Table 4. The dependent variable in all columns is PM2.5 measured in $\mu g/m^3$. We control for the number of COVID-19 cases and deaths at the national-level. The estimate for our countrywide lockdown is statistically significant at the 1% level and suggests that the introduction of a lockdown reduces PM2.5 levels by about 1.7 $\mu g/m^3$ in our preferred specification. The mean of the dependent variable for pre-lockdown days is about 10.7 $\mu g/m^3$, suggesting that European lockdowns decreased air pollution by about 16%.

We also document the effect of lockdowns for each country separately in Appendix Table A12. Our country-level controls are necessarily dropped for this analysis since we estimate the effect of a national lockdown for each of the countries separately. We find large and significant reductions in air pollution for all five countries.[18]

Last, in Appendix Figure A8, we present our leads and lags estimates for Europe, using the specification which decomposes the level of PM2.5 by the number of days before and since the national order. Following a lockdown, we see a general reduction in air pollution levels. This (admittedly messy) trend continues downward for approximately 15 days before reversing slowly and returning to pre-order levels around 75 days after the initial lockdown is released. This return to normal levels and its 'spiky' estimation likely reflects both underlying heterogeneity in strictness of orders and different country implementation and lifting of orders.

3.5 Safer-at-Home Orders and Collisions

We now present the main results for collision. Safer-at-home orders are implemented primarily to save lives by limiting the spread of the virus through social distancing, non-critical business closures, and restriction to only necessary activities. As a by-product, these orders can also result in fewer vehicles on the road– directly reducing air pollution, collisions, and perhaps even fatalities. As motor vehicle collisions are one of the leading causes of deaths for Americans, these unintended benefits would suggest that the number of lives saved by safer-at-home orders may be more than expected.[19]

[18] Of note, we document that Germany's partial order also led to a large decrease in air pollution, potentially suggesting that the stringency of lockdowns might not be related to the decrease in PM2.5 concentrations. See Brodeur et al., 2021 for more details on the stringency of European lockdowns.

[19] Deaths from motor vehicle collisions are surpassed only by heart disease, malignant neoplasms, and unintentional poisoning (both heart disease and malignant neoplasms have been connected to PM2.5 exposure).

3.5.1 Main Results

In Table 5, we estimate the effects of state orders on the collision incidence rate in a county, per day.[20] We present the incidence-rate ratios of a Poisson count model with county and date fixed effects; an estimate below one is a reduction in the dependent variable. The time period is January 1st, 2020 through the moment the statewide safer-at-home order ends.[21] The structure of the table is the same as Table 2.

In the first column, we estimate that a state order reduces the incidence of collisions by 16%.[22] In columns 2–3, we introduce the number of COVID-19 cases and deaths, respectively. Regardless of the underlying severity of the infection, the effectiveness of the state order remains large and statistically significant. In the fourth column, we introduce the set of other policies, while column 5 adds weather controls.[23] Introducing these additional variables in the model decreases the size of the coefficient (i.e., increases the decrease in car collision) and has no effect on the significance of the estimates. In column 5, the estimate suggest that safer-at-home orders decrease daily collisions by 20%. As there are approximately 1.4 collisions for treated counties per county and per day during the safer-at-home order, this reduction implies a counterfactual of about 1.7 collisions results in a reduction of 0.35 collisions per day *per county*.

In Appendix Table A14, we repeat the analysis presented in Appendix Table A2 and test whether counties that implemented an order prior to a statewide order are differently affected than counties that did not implement their own order. The estimate suggests that safer-at-home orders decrease car collisions for both sets of counties with a slightly larger decrease for counties that implemented an order prior to the statewide order. In addition, we find that conditional on the presence of a statewide order, county orders also statistically significantly decrease collisions. The estimated negative effect of a county order is equal to that of the statewide order, suggesting that counties also have the ability to significantly reduce collisions by issuing safer-at-home orders, even when under the influence of a statewide order.

To sum up, we find that state and county orders significantly decrease collisions. This result is quite important given the large number of car crash fatalities in the U.S. (about 35,000 in 2016).

3.5.2 Graphical Evidence

We provide a visual summary of the collision impact in Figure 4. This figure is similar to Figure 3 and plots the estimated collision levels at daily intervals pre- and post-statewide order (as in equation 2). Our estimates are not statistically significant from 40 days to a few days prior to the implementation of the order. There is a small increase in collisions a few days prior to the implementation of the order, perhaps due to growth in traffic for last minute shopping. Then we document a large decrease in collisions during safer-at-home orders, with the largest estimated coefficients from 45 to 80 days post-order. In other words, collisions are significantly

[20] See Appendix Table A13 for the analysis at the state-level rather than at the county-level. Our conclusions remain unchanged.

[21] The sample includes 1,711 counties. Counties with no collisions during the entire time period are excluded from the sample.

[22] Figure 4 confirms this pattern. This figure plots regression coefficients for collisions corresponding to number of days before/after state order issued.

[23] Precipitation increases daily collisions in our sample, while higher maximum temperature is associated with fewer collisions - unsurprising when the sample period assigns rising temperatures to coming out of winter rather than entering into the hottest months of summer.

reduced for as long as safer-at-home orders are in place. Once orders start to be lifted (i.e., all lifted after day *(t+80)*), the effect of the order on collision decreases in magnitude and becomes insignificant around 95 days after the order. These results suggest that orders do not have persistent effect on collisions.

3.5.3 Severity of Collisions

We now turn to the impact of safer-at-home orders on the severity of collisions. The severity of each collision is coded based on traffic flow/disruption and graded from one to four by the data providers. A value of one indicates a short delay as a result of the accident while a four indicates a significant impact on traffic, i.e., a long delay. Before the order period, the least common category of collision was the least severe. During the order period, the least common category was the most severe. During both periods, the most common severity of a collision is category two.

The estimates are presented in Table 6. All columns include our full set of controls. Column 1 reproduces our main results for any of the collision severities, while columns 2–5 look at the impact of state-wide orders on each of the four severity categories, respectively. The number of counties varies across columns since counties with no collisions in a given severity category are omitted.[24] Our estimates suggest that statewide safer-at-home orders significantly decreased collisions of severity one and two, by 14% and 23% respectively. There is no effect for the second-to-most severe crashes. Interestingly, we document increases in the most severe type of collisions, with an increase of 18% (a large percentage increase, however, this increase is relative to a small baseline). This result is in line with the idea that some drivers might be speeding more during lockdowns, which could lead to an increase in severe (and often fatal) collisions.

3.5.4 Strictness of Order and Heterogeneity Analyzes

We now check whether more stringent safer-at-home orders lead to a greater decrease in collisions than orders simply recommending to stay home. The estimates are presented in Appendix Table A9, columns 3 and 4. Again, our variables of interest are dummies for “low” safer-at-home intensity and “high” safer-at-home intensity. Our estimates show that both types of orders lead to a decrease in collisions of about the same magnitude.

Appendix Table A15 provides the heterogeneity analysis by county characteristics. The structure of this table is similar to Appendix Table A10. We find that the decrease in collisions is driven entirely by urban counties, with a significant decrease of about 23% in daily collisions. By political divide, we also find that counties that supported President Trump in 2016 have a smaller reduction in collisions than those who did not. By resident age, the difference seems to be small. We find that the decrease in collisions is very large and significant for counties in states with above median shares of occupations that can be done remotely, while the estimate is slightly greater than one (corresponding to an increase in collisions) for counties below the median shares.

To sum up, our results suggest that the documented decrease in collisions does not persist over time, is driven by urban counties, counties with lower support for President Trump, and counties in which workers are more able to work remotely.

[24]See Table 1 for summary statistics.

3.6 Stay-Home Orders, Social Distancing and Collisions

We now investigate one of the mechanisms through which safer-at-home policies might have impacted car collisions; changes in social distancing behaviors. For this analysis, we rely social distancing cell phone data from Unacast.

We proceed in three steps. First, we note that a large number of studies have documented the impacts of safer-at-home on social mobility, including our own working paper Brodeur, Cook, and Wright, 2020. See, for instance, Brodeur, Grigoryeva, and Kattan, 2020 and Cicala et al., 2020 who rely on Unacast data and provide evidence that safer-at-home policies decreased total distance traveled.[25]

Second, we test whether safer-at-home orders changed the daily distribution of car collisions. In other words, we want to explore whether or not the change in collisions caused by safer-at-home orders is partly coming from a change in the timing of travel (and therefore a change in the relative congestion levels of different times of day) rather than entirely due to reductions in travel. In addition this may have implications for the composition of collisions in terms of single or multiple vehicles involved. Figure 5 illustrates the distribution of collisions across all hours of the day for our sample period and the corresponding time period in 2019. We find that the timing of collisions changed in 2020 in comparison to 2019, with more collisions in 2020 occurring during the afternoon and less during the night and early rush hour. A large literature documents a positive or concave relationship between traffic/congestion and car crashes (e.g., Gwynn, 1967; Head, 1959; Schoppert, 1957; Zhou and Sisiopiku, 1997). See Retallack and Ostendorf, 2019 for a literature review. It is posited that this relationship could be a result of aggregating single vehicle and multi-vehicle collisions where single vehicle collisions are high during periods of little congestion and multi-vehicle collisions are high during periods of high congestion. As safer-at-home orders increase work from home and decrease mobility, we should expect to see less congestion and as a consequence fewer collisions during the typical rush hour periods. Similarly, as individuals now have more flexibility in their schedules, we might expect trips to be displaced across hours of the day rather than eliminated entirely and as a consequence we should expect an increase in fatal collisions. While we do not observe collision fatality directly, we do provide evidence of increased collision severity as measured by traffic delays in Table 6. Overall, these findings provide suggestive evidence that the increase in more severe collisions during safer-at-home orders is partly due to a shift in the distribution of traffic (and therefore congestion) across hours of the day.

Third, we document the relationship between travel distance and collisions by exploiting large variation in mobility due to safer-at-home orders. As the typical structure of work days and commutes was significantly altered by safer-at-home orders, it is possible that the positive or concave relationship between congestion/traffic and collisions no longer holds while they are in effect. While social mobility data does not provide direct evidence of congestion, we can use this data to explore the suggestive evidence of the relationship between travel and collisions using the exogenous variation in traffic from safer-at-home orders. However, we are unable to say anything about the concave nature of the relationship using our empirical strategy.

To attempt to achieve exogenous variation of mobility at the county-level, we instrument travel distance with statewide safer-at-home orders. The rationale for

[25] This relationship has also been shown by Google back in April 2020 (see https://www.theverge.com/2020/4/3/21206318/google-location-data-mobility-reports-covid-19-privacy).

the instrument is that statewide orders led to a large decrease in mobility, which we exploit to document the relationship between mobility and collisions.

More precisely, we estimate:

$$\begin{cases} Travel_{cst} = \rho + \phi \cdot StateSafer_{st} + X'_{cst}\psi + \theta_c + \kappa_t + \nu_{cst} \\ Y_{cst} = \alpha + \delta \hat{Travel}_{cst} + X'_c\gamma + \theta_c + \kappa_t + \varepsilon_{cst}, \end{cases} \tag{3}$$

where $StateSafer_{st}$ equals one once the state has implemented the order and zero for the pre-policy period. We run a first stage in which we regress this variable on the travel distance at the county-level, including all controls and fixed effects as in Equation 1. Then we plug in the predicted values of this first stage and estimate the second stage of the 2SLS. The dependent variable in the second stage is the number of traffic collisions. The time period is March 1st, 2020 through the moment the statewide safer-at-home ended.

For our instrument to be valid two conditions have to hold. First, our instrument has to be a strong predictor of travel distance. As mentioned before, a large literature showed that statewide safer-at-home orders significantly decreased mobility. The F-statistic for the first stage is about 1733 confirming the strong negative impact of orders on mobility.

Second, in order for our instrument to allow a causal interpretation, statewide orders must only affect the number of collisions through its effect on social mobility, i.e., the exogeneity assumption. We believe this condition is unlikely to hold in our setting given the documented effect of lockdowns on economic activity and other socioeconomic variables. This is an issue in our setting since it is plausible that labor force status and work arrangements are related to driving behavior. Nonetheless, we proceed with our 2SLS exercise, but caution readers that the exclusion restriction is likely violated.

Table 7 presents our estimates of Equation 3. Column 1 is for all counties, column 2 corresponds to majority urban counties, and column 3 to majority rural counties. The estimate in column 1 suggests that a 100% increase in travel distance relative to the baseline period is associated with 4 additional car collisions per county-day. The estimate is statistically significant at conventional levels. The estimate in column 2 is also statistically significant and suggests a 100% increase in travel distance relative to the baseline period is associated with just over 9 additional collisions per county-day in urban counties. Meanwhile, we find no evidence that changes in travel distance induced by safer-at-home orders reduce collisions in rural counties. Taken together, these findings provide suggestive evidence that the reduction in collisions stemming from safer-at-home orders is being modulated through mobility and is driven by urban counties rather than rural counties. Consistent with the existing literature, the positive relationship between congestion (proxied for by travel distance) and collisions remains during the COVID-19.

To summarize, we argue that safer-at-home orders led to a decrease in social mobility and a shift in the timing of traffic away from traditionally congested periods, and show suggestive evidence that these are two of the main driving forces behind the observed reduction in collisions and increase is collision severity.

3.7 Interpretation

In this section, we provide back-of-the-envelope calculations of the benefits from *positive* pollution and collision externalities generated by safer-at-home orders. Our calculations here are, in part, based on the growing literature estimating the revealed

- rather than stated - willingness to pay (WTP) for air quality. It is also important to note that most studies do not provide estimates of the WTP for temporary reductions in air quality, but instead attempt to identify the WTP of a permanent reduction in pollution. This means that our calculations are based on the assumption that individuals value a temporary abatement at the average annual value of a permanent reduction pro-rated for the duration of the abatement period, in this case the duration of the safer-at-home order. Note that these calculations ignore possible WTP increases for clean air during COVID-19, as recent research has begun to identify increased mortality from and transmission of the virus with higher levels of contemporaneous air pollution (Zhang, Xue, and Jin, 2020; Pansini and Fornacca, 2020).

We first review the literature. Currie et al., 2015 exploit the effect of toxic plant openings to estimate the impact of air quality on American house values and birth weights. They find an 11% reduction in house values and a 3% increase in low birth weights in nearby households. Recent work by Ito and Zhang, 2020 using purchases of air purifiers in China suggests that the mean WTP for 1 $\mu g/m^3$ reduction of PM10 is 1.34 USD annually - and increases strongly with household income. For a household making 10,000 USD per year (the upper limit of the sample), they estimate a marginal willingness to pay of 5 USD per 1 $\mu g/m^3$ PM10. Chay and Greenstone, 2005 study housing price evolutions in the 1970's and '80s in American counties that were quasi-randomly assigned federally mandated air pollution regulations. They find that a one unit decrease in particulates (all suspended particulates - what was targeted by the regulations) results in a 0.7 to 1.5 percent increase in house values. Deschênes, Greenstone, and Shapiro, 2017 quantify the defensive investment portion of willingness to pay for air pollution reduction to be around one third - and that nitrogen oxide reduction program benefits 'easily' exceed costs. Barwick et al., 2017 estimate that the lower bound of the annual WTP for a 10 $\mu g/m^3$ reduction in PM2.5 is 9.25 USD per Chinese household, or 7% of total healthcare spending. Finally, Bayer et al., 2016 estimate the WTP to avoid ozone using house purchases in the San Francisco Bay Area, finding a 10% reduction in pollution commanded a price almost equal to a 10% reduction in violent crime.

We now turn to our back-of-the-envelope calculations. To calculate these, we first rely on estimates from the U.S. to compute the WTP associated with our estimated 7%–25% reduction in PM2.5. We then scale the estimates to the duration of each state's safer-at-home order and aggregate over the number of households in the state (drawn from the 2018 American Community Survey) before finally aggregating WTP over all states implementing policies.

Recall that our estimates from Table 2 indicate that the introduction of safer-at-home orders decreased pollution by about 1.4 $\mu g/m^3$ while Appendix Table A8 provides an estimate of 0.4 $\mu g/m^3$. We also note that Bayer et al., 2016 find that American home owners are willing to pay between about 300 USD annually for a 10% reduction in one pollutant; the WTP associated with our estimated 7%–25% reduction in PM2.5 could thus be as high as 210–750 USD annually per household.[26] Using WTP estimates from the most appropriate American samples, we find estimated benefits of 154 million to 500 million USD using the adapted WTPs from Bayer, Keohane, and Timmins, 2009 (who estimate that the marginal WTP for an annual 1 $\mu g/m^3$ reduction in PM10 for United States metropolitan areas to be 22 USD per household), and 3.6 billion to 13.1 billion USD using the adapted estimates

[26] While this may seem large, classical estimates from Harrison and Rubinfeld, 1978 using data from the Boston Area following the *Clean Air Act* placed a WTP for a 25% reduction in air pollution at approximately 2,000 (1978 USD). Other WTP estimates discussed in Chattopadhyay, 1999 for particulate pollution reductions in the Chicago Area up to 366 USD in 1982-84 dollars.

from Bayer et al., 2016. These estimates vary widely, no doubt due to the many assumptions necessary to compute figures at the aggregate level. However, they serve to give a sense of the order of magnitude of the possible environmental benefits these orders have.

There are also extensive costs associated with traffic collisions, from congestion impacts; to medical and repair bills; to loss of life. We generate rough estimates of the benefits of reduced collisions using the National Highway Traffic Safety Administration's (NHTSA) estimates that the average cost of collisions in 2013 was 17,794 USD (2010, 21,054 USD in 2020) per crash.[27] When accounting for quality-of-life valuations, the estimates are an average of 61,470 USD (2010, 72,732 USD in 2020) per crash.

Our estimates in Table 5 suggest a reduction in collisions of 20%. The mean number of crashes during safer-at-home orders among treated counties was 1.4 crashes per day, implying that the counterfactual crash rate would have been about 1.77 crashes per day per county. Applying our estimates to the 124,370 county-days spent under safer-at-home policies suggests that over 219,000 collisions may have been avoided by June 30th, 2020. Using the numbers from the NHTSA gives approximately 15.9 billion 2020 USD in costs avoided as a result of safer-at-home orders over that period.

A major limitation is that our estimates in Table 6 suggest that we actually see an increase in more severe collisions while the average effects are driven by reductions in less severe collisions. This means that our back-of-the-envelope calculations may overestimate the costs of collisions as while the volume of crashes is declining, the fatality rate may not be. Unfortunately, our data do not contain information about fatalities directly and therefore do not allow us examine the extent to which this could be true.

3.8 Conclusion

In many respects, safer-at-home policies have been expected and shown to have negative impacts on societies by, for instance increasing mental health distress and exacerbating the economic impacts of COVID-19. This paper represents a first step toward understanding some of the unintended *positive* effects safer-at-home policies have on pollution and car crashes.

We rely on a difference-in-differences framework with high frequency air pollution data and daily collision data. We find that statewide safer-at-home policies lead to a 25% reduction in PM2.5 concentrations and a 20% reduction in vehicular collisions; one of the leading causes of death in the United States. We also provide suggestive evidence that the reduction in collisions is driven in part by reduced travel associated with safer-at-home orders and by distributional changes to traffic times that also explain the increase is collision severity. We calculate that over 219,000 collisions may have been avoided by June 30th, 2020, which translates to approximately $16 billion in costs avoided. The benefits from reduced air pollution could range from $154 million to $13 billion.

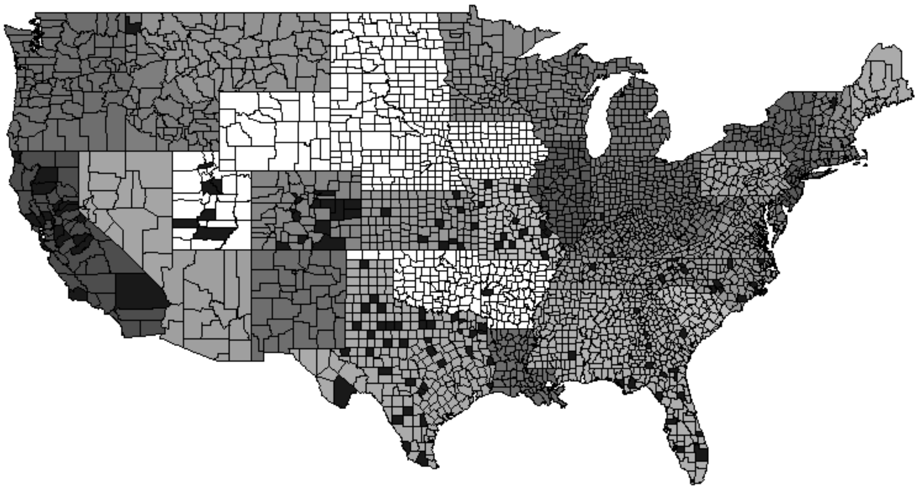
Our paper raises broader questions of the nuance involved in estimating the costs and benefits of safer-at-home policies. As more data on COVID-19 cases and deaths became available, it was possible to better estimate how many lives were saved (Hsiang et al., 2020). But the unintended economic consequences and large sphere of

[27] See https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013.

domains impacted by safer-at-home orders make it a difficult, but worthwhile, task to estimate the full set of costs and benefits of these policies.

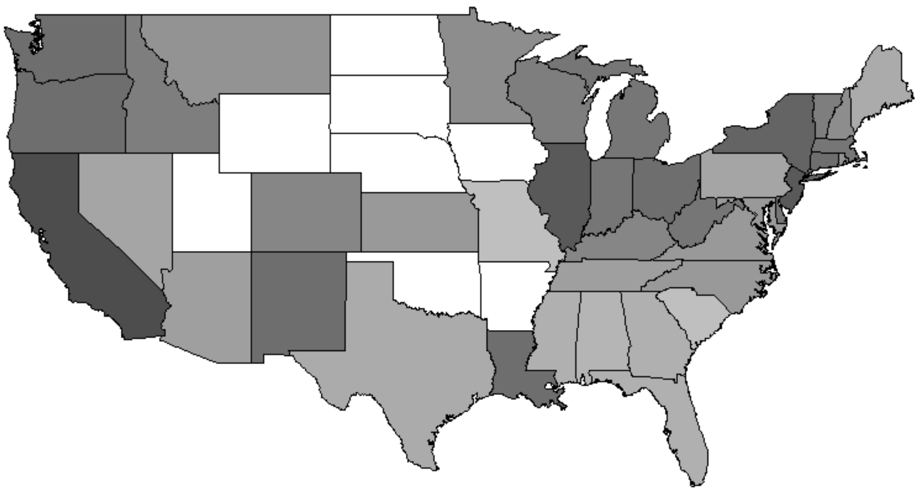
3.9 Figures

FIGURE 1: Counties that Issued an Order



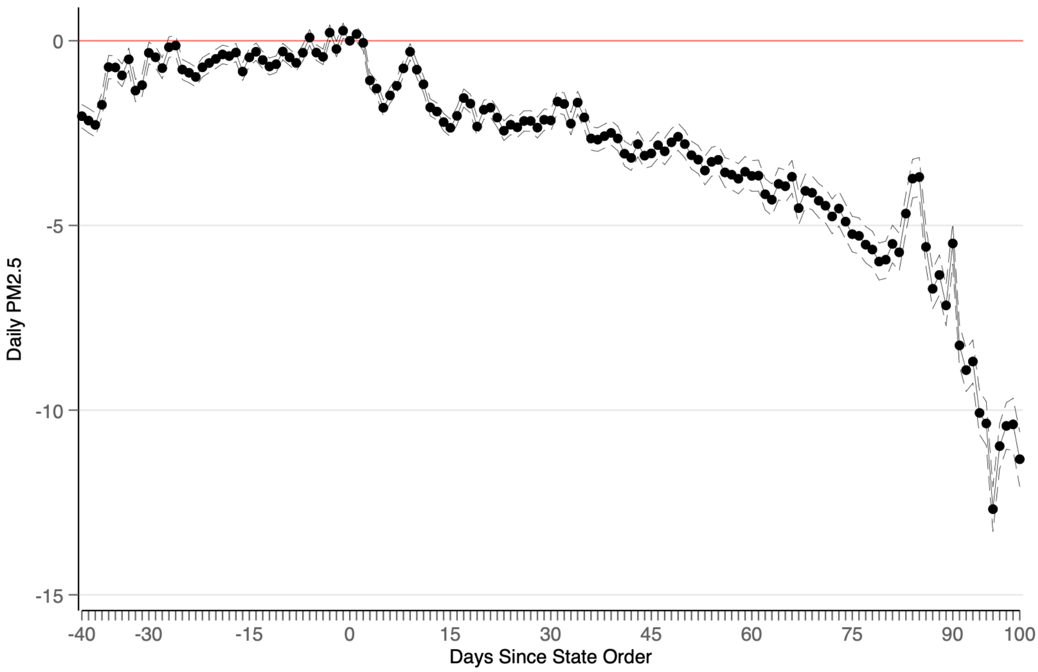
Notes: This map presents counties and states that issued an order prior to April 30, 2020. Counties that issued their own order prior to their state shaded darkest. For states, the darker the fill, the earlier the state issued the order. States in white did not issue an order.

FIGURE 2: States that Issued a Lockdown



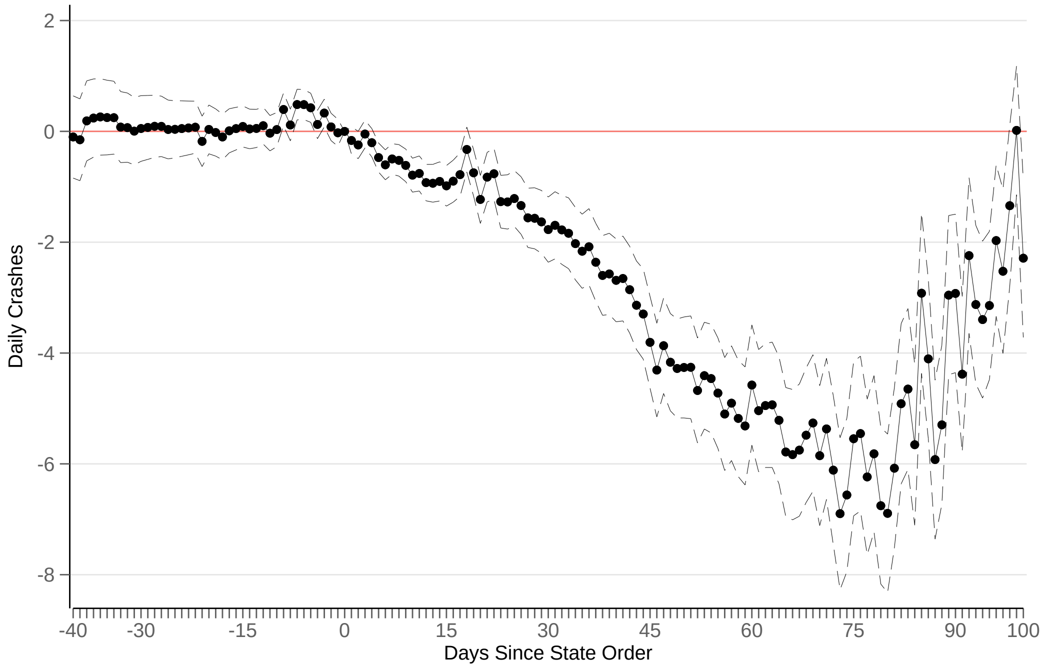
Notes: This map presents states that issued an order prior to April 30, 2020. The darker the fill, the earlier the state issued the order. States in white did not issue an order.

FIGURE 3: PM2.5 Concentrations Over Time



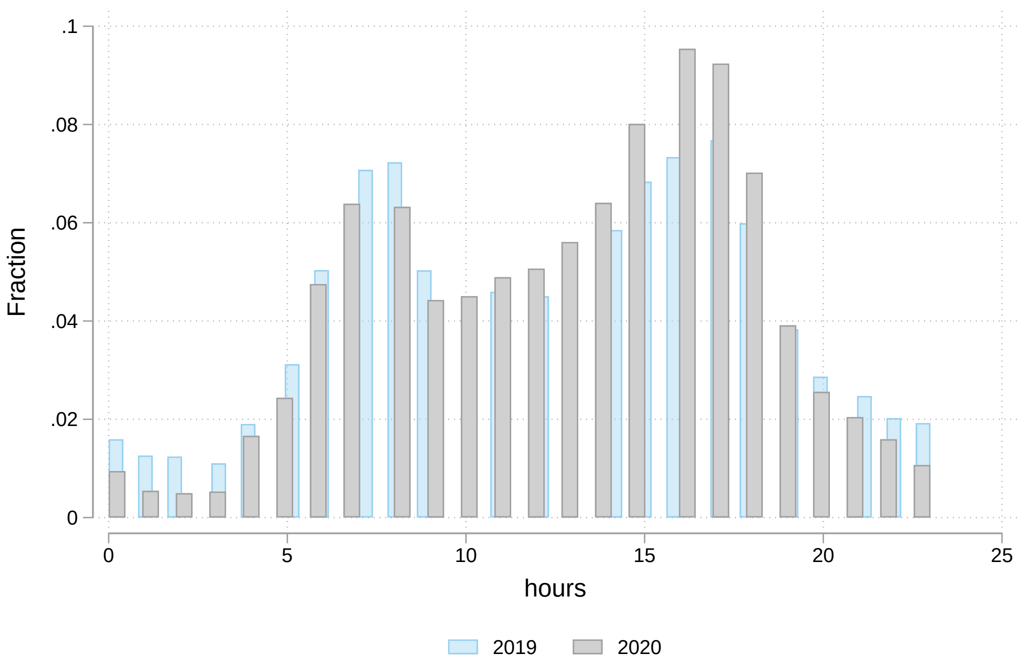
Notes: This figure presents regression coefficients for PM2.5 concentrations corresponding to number of days before/after state order issued. Temperature, precipitation, COVID-19 cases and deaths, and other policy controls are included along with date and county fixed effects. Confidence intervals at 95% presented.

FIGURE 4: Traffic Collisions Over Time



Notes: This figure presents regression coefficients for collisions corresponding to number of days before/after state order issued. Temperature, precipitation, COVID-19 cases and deaths, and other policy controls are included along with date and county fixed effects. Confidence intervals at 95% presented.

FIGURE 5: Traffic Collisions Across Hours of Day



Notes: This figure presents histograms for collisions across all hours of the day for our sample period and the corresponding time period in 2019.

3.10 Tables

TABLE 1: Summary Statistics

	Mean	Std.Dev.	Max	Min	Count
Pollution					
PM2.5	6.641	3.534	84.4	0.1	200094
PM2.5 (UK)	9.977	8.015	45.0	1.4	25906
PM2.5 (GE)	8.691	6.427	85.4	.7	66407
PM2.5 (SP)	6.984	4.533	34.9	1.0	9648
PM2.5 (FR)	8.238	5.803	189.1	0.5	23880
PM2.5 (IT)	8.594	4.559	70.6	1.4	29487
Collisions					
Collisions	1.770	9.949	811.0	0.0	237408
Severity 1 collisions	0.081	1.095	72.0	0.0	237408
Severity 2 collisions	1.293	8.301	681.0	0.0	237408
Severity 3 collisions	0.334	1.844	116.0	0.0	237408
Severity 4 collisions	0.061	0.537	31.0	0.0	237408
COVID-19					
COVID cases per 10k	5.946	22.448	625.5	0.0	237408
COVID deaths per 10k	0.264	1.262	32.0	0.0	237408
County-days under lockdown	0.364	0.481	1.0	0.0	237408
Day care closure	0.125	0.330	1.0	0.0	237408
Eviction moratorium	0.223	0.416	1.0	0.0	237408
Mandatory face mask in public	0.055	0.229	1.0	0.0	237408
Mandatory quarantine for visitors	0.019	0.138	1.0	0.0	237408

Notes: Authors' calculations. PM2.5 is 24-hour daily concentration of PM2.5 in $\mu g/m^3$. Collisions data gathered from Moosavi et al., 2019. The severity of an accident is coded as a number ranging from 1 to 4 with 1 being the smallest impact on traffic and 4 being a significant impact on traffic.

TABLE 2: State Orders and Pollution (PM2.5)

	(1)	(2)	(3)	(4)	(5)
	PM2.5	PM2.5	PM2.5	PM2.5	PM2.5
During safer-at-home-order	-1.671	-1.673	-1.673	-1.607	-1.372
	(0.454)	(0.454)	(0.455)	(0.496)	(0.413)
COVID cases per 10k		-0.003	-0.003	-0.003	-0.002
		(0.004)	(0.004)	(0.004)	(0.004)
COVID deaths per 10k			0.003	-0.004	-0.006
			(0.033)	(0.034)	(0.035)
Constant	6.704	6.704	6.704	6.709	6.460
	(0.809)	(0.808)	(0.808)	(0.816)	(0.801)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	200094	200094	200094	200094	200094
Counties	1592	1592	1592	1592	1592

Notes: State orders significantly reduce PM2.5. The dependent variable is average daily PM2.5 concentration at the county-level. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects. Sample restricted to counties within 50km of an air pollution monitoring station.

TABLE 3: State Orders and Polluted Days

	(1)	(2)	(3)	(4)	(5)
	PM2.5>12	PM2.5>12	PM2.5>12	PM2.5>12	PM2.5>12
During safer-at-home-order	-0.103	-0.103	-0.104	-0.099	-0.089
	(0.032)	(0.032)	(0.032)	(0.035)	(0.031)
COVID cases per 10k		-0.000	-0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
COVID deaths per 10k			-0.001	-0.001	-0.002
			(0.002)	(0.002)	(0.002)
Constant	0.109	0.109	0.109	0.109	0.093
	(0.039)	(0.039)	(0.039)	(0.039)	(0.038)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	200094	200094	200094	200094	200094
Counties	1592	1592	1592	1592	1592

Notes: State orders significantly reduce polluted days. The dependent variable takes a value of 1 if PM2.5 is above the Annual National Ambient Air Quality Standard of 12 $\mu g/m^3$. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 100 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects.

TABLE 4: European National Orders and Pollution (PM2.5)

	(1) PM2.5	(2) PM2.5	(3) PM2.5	(4) PM2.5	(5) PM2.5
During lockdown	-1.047***	-1.069***	-0.969***	-1.676***	-1.690***
	(0.036)	(0.033)	(0.036)	(0.037)	(0.037)
COVID cases per 10k		-0.005***	-0.003**	0.003**	0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
COVID deaths per 10k			-0.117***	-0.015	0.034**
			(0.015)	(0.016)	(0.016)
Constant	35.373***	24.137***	24.151***	24.254***	24.055***
	(0.135)	(0.121)	(0.121)	(0.119)	(0.119)
Date FE	Y	Y	Y	Y	Y
Mask mandate				Y	Y
Quarantine mandate				Y	Y
School closures				Y	Y
Temperature					Y
Precipitation					Y
Observations	155328	138568	138568	138568	136300

Notes: National orders significantly reduce PM2.5 in France, Germany, Italy, Spain and the United Kingdom. The dependent variable is average daily PM2.5 concentration, measured at the population centroid for an administrative area from the three nearest air quality monitors. The mean concentration during the period is 8.7, and the within-unit standard deviation is 6.0. An observation is an area-day. A total of 841 areas are used with 96 from France (departments), 403 from Germany (kreise), 113 from Italy (provinces), 55 from Spain (provinces) and 192 from the United Kingdom (counties). Observations are from the first to the 300th day of 2020, which includes both before, during, and after lockdowns. Standard errors reported in parentheses. All columns include area and date fixed effects.

TABLE 5: State Orders and Collisions

	(1)	(2)	(3)	(4)	(5)
During safer-at-home-order	0.840	0.839	0.838	0.826	0.793
	(0.050)	(0.050)	(0.050)	(0.045)	(0.045)
COVID cases per 10k		1.002	1.003	1.002	1.002
		(0.000)	(0.001)	(0.001)	(0.001)
COVID deaths per 10k			0.986	0.981	0.983
			(0.012)	(0.011)	(0.011)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	237569	237569	237569	237569	237408
Counties	1711	1711	1711	1711	1710

Notes: State orders significantly reduce traffic collisions. Poisson model with fixed effects. The dependent variable is count of traffic collisions at the county-level. Coefficients are incidence rate-ratios, wherein a value below one indicates a decrease in the dependent variable and a value above indicates an increase. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 100 people. Robust standard errors reported in parentheses. All columns include county and date fixed effects.

TABLE 6: State Orders and Collision Severity

	(1) Any	(2) Severity 1	(3) Severity 2	(4) Severity 3	(5) Severity 4
During safer-at-home-order	0.793	0.855	0.769	1.006	1.179
	(0.045)	(0.031)	(0.034)	(0.087)	(0.146)
COVID cases per 10k	.998	0.998	1.001	1.004	1.004
	(0.001)	(.001)	(0.001)	(0.001)	(0.001)
COVID deaths per 10k	1.033	1.033	0.969	0.970	1.013
	(0.011)	(.0179)	(0.017)	(0.012)	(0.016)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies	Y	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y	Y
Observations	237408	60138	207757	166441	145069
Counties	1710	418	1490	1189	1031

Notes: Poisson model with fixed effects. The dependent variable is count of traffic collisions at the county-level. Coefficients are incidence rate-ratios, wherein a value below one indicates a decrease in the dependent variable and a value above indicates an increase. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 100 people. Robust standard errors reported in parentheses for all columns except column 2, which is non-singular when robust standard errors are applied. All columns include county and date fixed effects.

TABLE 7: Travel Distance and Collisions – Instrumental Variable

	(1) All	(2) Urban	(3) Rural
Travel Distance	4.061	9.187	0.071
	(0.999)	(2.576)	(0.133)
County FE	Y	Y	Y
Date FE	Y	Y	Y
Case & Death rates	Y	Y	Y
COVID-19 policies	Y	Y	Y
Weather controls	Y	Y	Y
Kleibergen-Papp F-stat	1733	807	755
Observations	159488	159488	228352

Notes: 2SLS with fixed effects. The dependent variable (second stage) is the number of traffic collisions at the county-level. The instrumental variable is the presence of a statewide safer-at-home policy. The time period is March 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 100 people. Robust standard errors reported in parentheses. All columns include county and date fixed effects.

3.11 Appendices

Appendices

3.11.1 Non-Pharmaceutical Interventions

Day care closure: "The date a state closed daycares statewide. Only included directives/orders. Did not include guidance or recommendations. Order must apply to entire state."

Face masks: "The date a state mandated face mask use in public spaces by all individuals statewide. The order does not have to apply to all public spaces, but must apply state wide. Only included directives/orders. Did not include guidance or recommendations. Order must apply to entire state."

Freezes on evictions: "The date a state stopped the initiation of evictions (overall or due to COVID-19 related issues) statewide. This could be mandated from governors or though the state court system. Did not include guidance or recommendations. Order must apply to entire state."

Quarantine: "The date a state first mandated that individuals arriving in their state from any state must undergo quarantine. Did not include guidance or recommendations. Order must apply to entire state. Quarantine order must apply to visitors using all forms of transportation to enter the state (not just air travel)."

3.11.2 Simulations Using Synthetic Control Methods: Pollution

Simulations Methodology

Here we detail a robustness check for the magnitude of our findings and construct a slightly different comparison group.

In our setting, areas can be one of three groups at any point in time, a) areas that have not *yet* had a state order applied to them b) areas that will *never* have a state order applied to them, or c) areas that currently have a state order applied. In an effort to construct a counterfactual (what would PM2.5 concentrations looked like in areas that currently have a state order applied), we use synthetic control methods to form a weighted convex combination of areas that never have a state order applied for every (eventually) treated area.

The estimates in this paper that use panel fixed effects and within-unit variation of the treatment variable use as a comparison group areas that are not-yet treated. This application of the Synthetic Control Methods (SCM) allows us to bring in information (in the form of a synthetic control area for each treated area) of the PM2.5 evolution that occurs in states during the treatment period. The SCM requires statistical 'donor' areas - which we restrict to be areas that are not ever treated - to donate their PM2.5 evolutions to those that are treated. We fit the SCM counterfactual on their pre-safer-at-home order PM2.5 average concentrations.

To construct our synthetic matching, we follow the steps in Abadie, Diamond, and Hainmueller, 2010. The idea of this method is to match a treated area (with a state order) to a group of control areas having similar pollution levels prior to the SAH order's implementation. The hypothesis is that the treated and control counties would have a similar change in pollution if the order had not been implemented.

We construct a synthetic match for each of the treated areas by solving the following optimization problem and finding the optimal vector of weights:

$$\forall i \in \mathbb{N}, \left\{w_j^{i*}\right\}_{j \in \mathbb{U}} = \arg\min \left\{w_j^i\right\} \sum_i \sum_t \left[Y_{it} - \sum_j w_j^i Y_{jt}\right]^2$$

Subject to

$$\sum_j w_j^i = 1 \quad \text{and} \quad \forall j \in \mathbb{U}, \quad \forall i \in \mathbb{N}, \quad w_j^i \geq 0,$$

Where Y_{it} is the pollution for county $i \in \mathbb{N}$ on pre-event $t \in \mathbb{T}$. $\mathbb{N}$ being the set of treated areas and $\mathbb{T}$ the set of pre-order dates. w_j^i is the weight given to county $j \in \mathbb{U}$, the set of control areas.

The pollution level for each synthetic area is constructed as:

$$\widehat{Y}_{it} = \sum_j w_j^{i*} Y_{jt}$$

The estimates are put together in the same manner as regressions elsewhere in the paper. Each area is equally weighted in the regressions.

Simulation Results

The estimates presented in this paper indicate a large reduction in PM2.5 after a state issues a safer-at-home order. As a robustness check, we use simulated counterfactual PM2.5 levels for each county in a state that issued an order using a 'synthetic' county constructed using the `synth` package (described in Abadie, Diamond, and Hainmueller, 2011). The results of this exercise are presented in Appendix Table A16 while those of a falsification exercise are presented in Appendix Table A17. The structure of the tables is the same as the bulk of those throughout the paper. Inference using simulations with these counterfactuals as we construct them is not explicitly endorsed by Abadie, Diamond, and Hainmueller, 2010, so we take these results as a robustness exercise, which provide reassurance that our results are, at the very least, not simply a statistical artifact. For this reason, we remain silent on the statistical significance of these estimates but provide them for the curious reader.

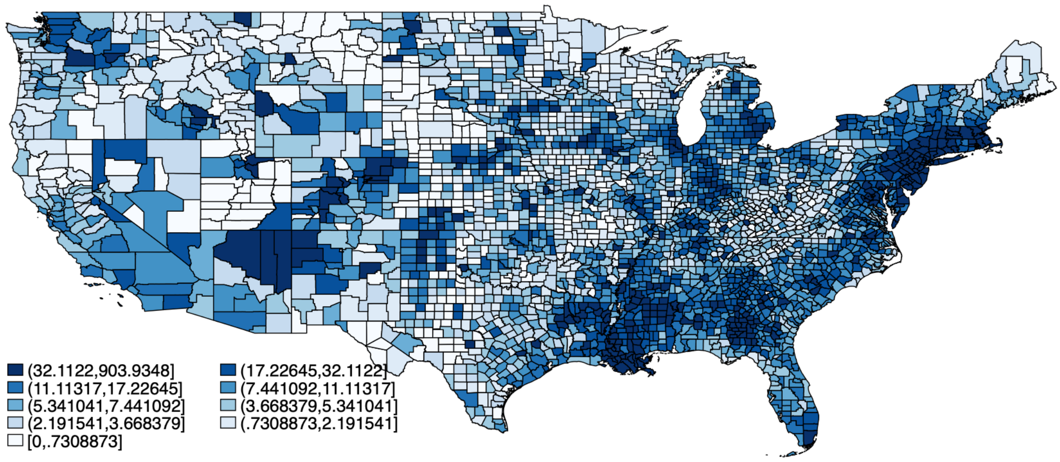
In Appendix Table A16, the dependent variable is the difference between the observed PM2.5 levels and those of the counterfactual. The difference is structured so that a negative value represents a lower observed PM2.5 concentration than expected from the counterfactual. To create the counterfactual, we applied the synthetic control method separately for each treated county, allowing for different donor (untreated) county weights to be applied to treated counties as diverse as those from New York or from Texas. The treated sample is necessarily restricted to 1418 counties in the 41 states that issued a safer-at-home order during 2020.

During a safer-at-home order, we see a reduction of around 1 μg/m^3 in PM2.5 concentration, which is reassuringly close to the estimates presented in Table 2. The estimates of the reduction are also relatively undisturbed with the introduction of controls including COVID-19 policies (such as mandatory mask usage) and weather.

In Appendix Table A17, we conduct a placebo-in-space exercise, much in the spirit of those introduced in Abadie, Diamond, and Hainmueller, 2010. The dependent variable is the difference between the observed PM2.5 levels of counties in states that did not issue a safer-at-home order in 2020 and those of a counterfactual constructed from counties in the 8 untreated states. To create the counterfactual, we applied the synthetic control method separately for each untreated county, allowing for different donor (untreated) county weights to be applied to the current untreated county. The sample is necessarily restricted to counties in the 8 states that issue a safer-at-home order during 2020. Reassuringly, this placebo exercise (which assigns a safer-at-home order when one never occurred) exhibits different behaviour than our simulation. The sign of the estimates even flips between specifications, with the inclusion of weather controls. The standard errors are significantly larger, and calculated in the same manner as in the simulation.

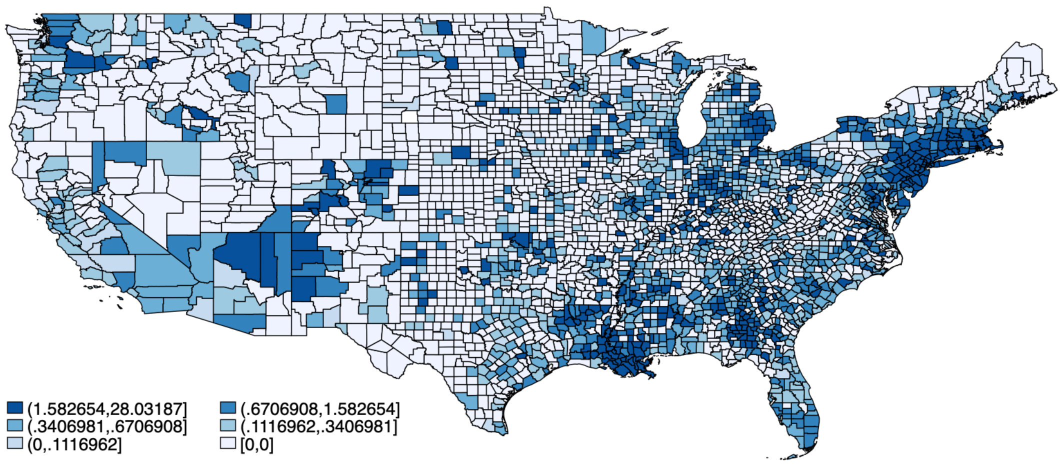
We also provide the same graphical evidence of the leads and lags for this and the placebo exercise. In Appendix Figure A9, we present the difference between observed and counterfactual PM2.5 levels from 40 days before to 100 days after a state order was issued. The results are similar to those presented in Figure 3. In contrast, the estimates for the placebo exercise presented in Figure A10 are much smaller and exhibit a flat trajectory before and after the placebo order's issuance.

FIGURE A1: COVID-19 Confirmed Cases per 10,000 by County



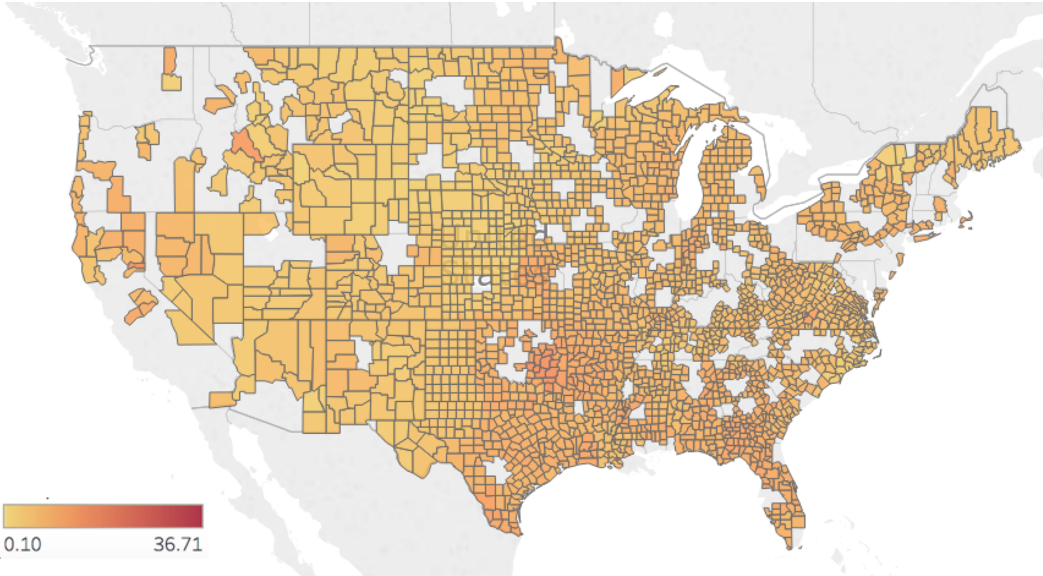
Notes: The map illustrates the cumulative number of (confirmed) COVID-19 cases per 10,000 inhabitants for each county as of May 1, 2020.

FIGURE A2: COVID-19 Deaths per 10,000 by County



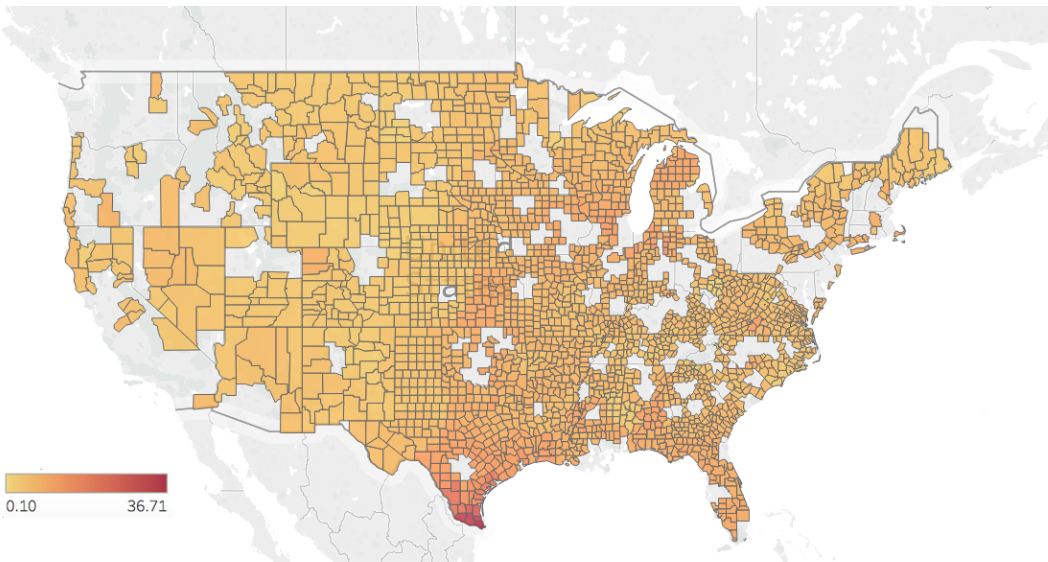
Notes: The map illustrates the number of COVID-19 deaths per 10,000 inhabitants for each county as of May 1, 2020.

FIGURE A3: Weekly Average PM2.5 Concentrations March 1-7, 2020



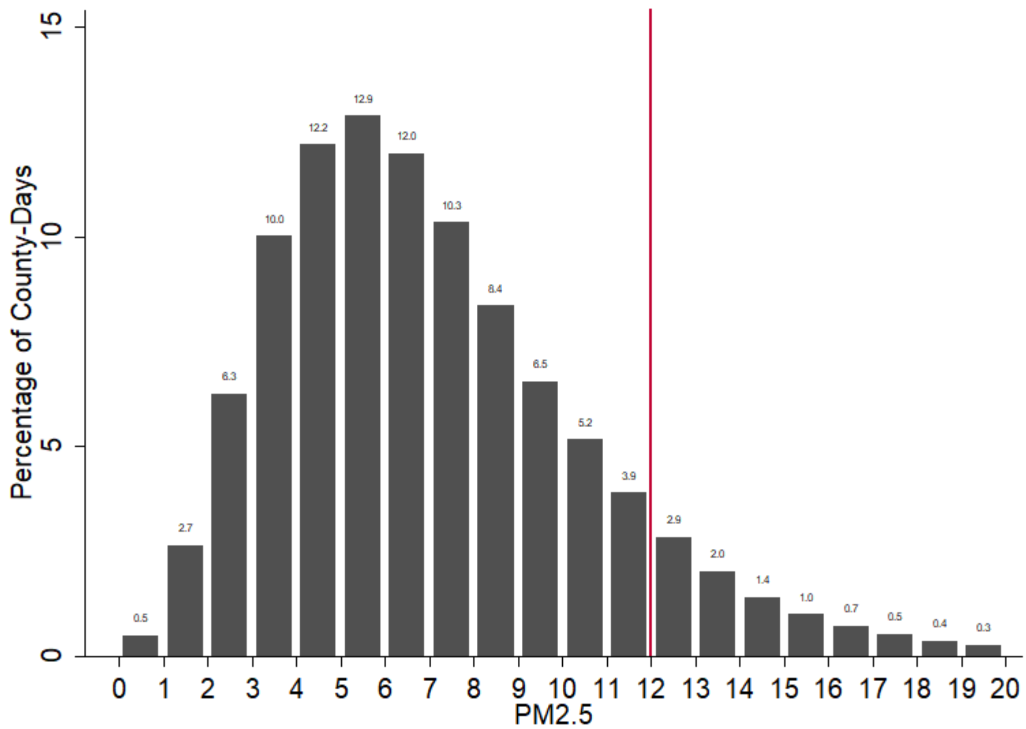
Notes: This map presents PM2.5 concentrations for a baseline week of March 1-7. PM2.5 measures for counties with a population weighted centroid within 50km of an air quality monitoring station.

FIGURE A4: Weekly Average PM2.5 Concentrations April 19-25, 2020



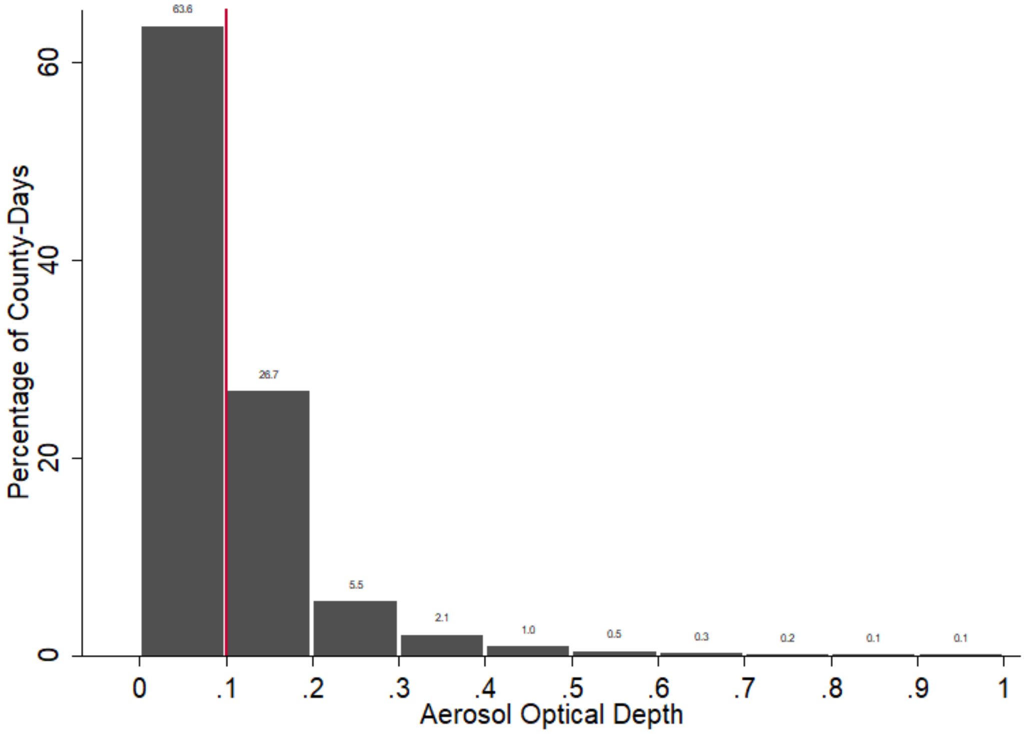
Notes: This map presents PM2.5 concentrations for near end of April 2020. PM2.5 measures for counties with a population weighted centroid within 50km of an air quality monitoring station.

FIGURE A5: Distribution of PM2.5 Concentrations



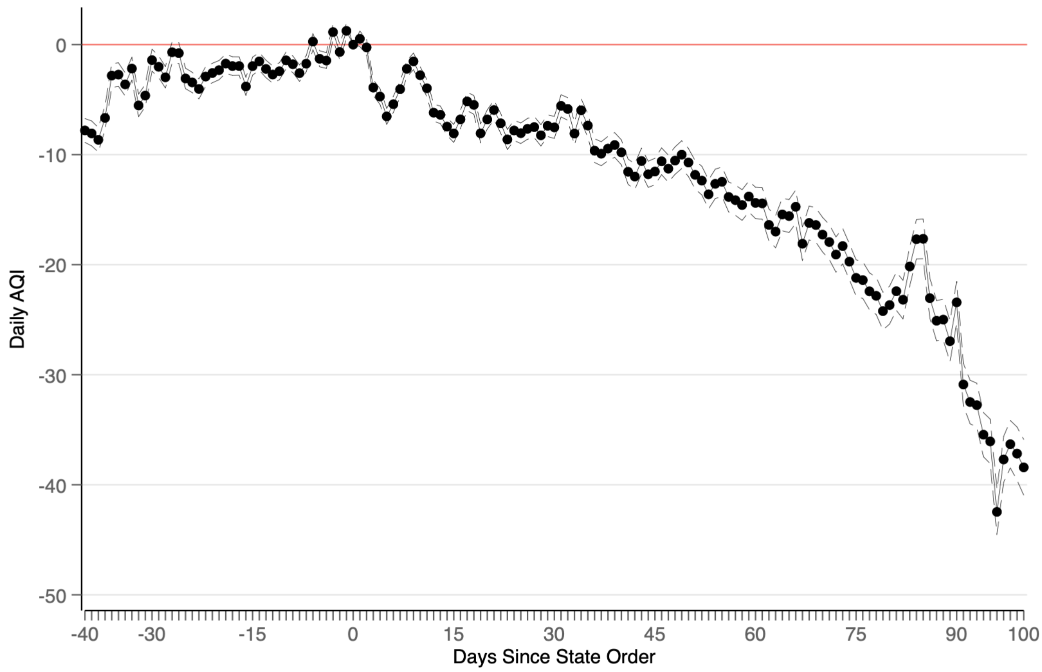
Notes: This figure illustrates the distribution of our dependent variable PM2.5 Concentrations for all county-days in our sample, i.e., all observations.

FIGURE A6: Distribution of Aerosol Optical Depth



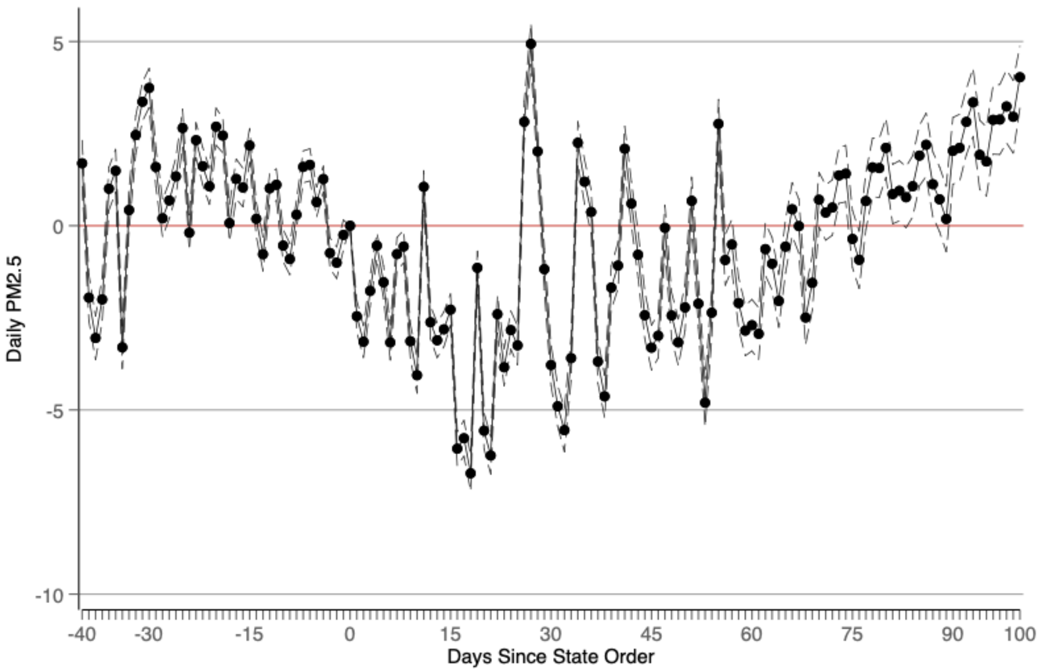
Notes: This figure illustrates the distribution of our dependent variable aerosol optical depth for all county-days in our sample, i.e., all observations.

FIGURE A7: Air Quality Index (AQI) Over Time



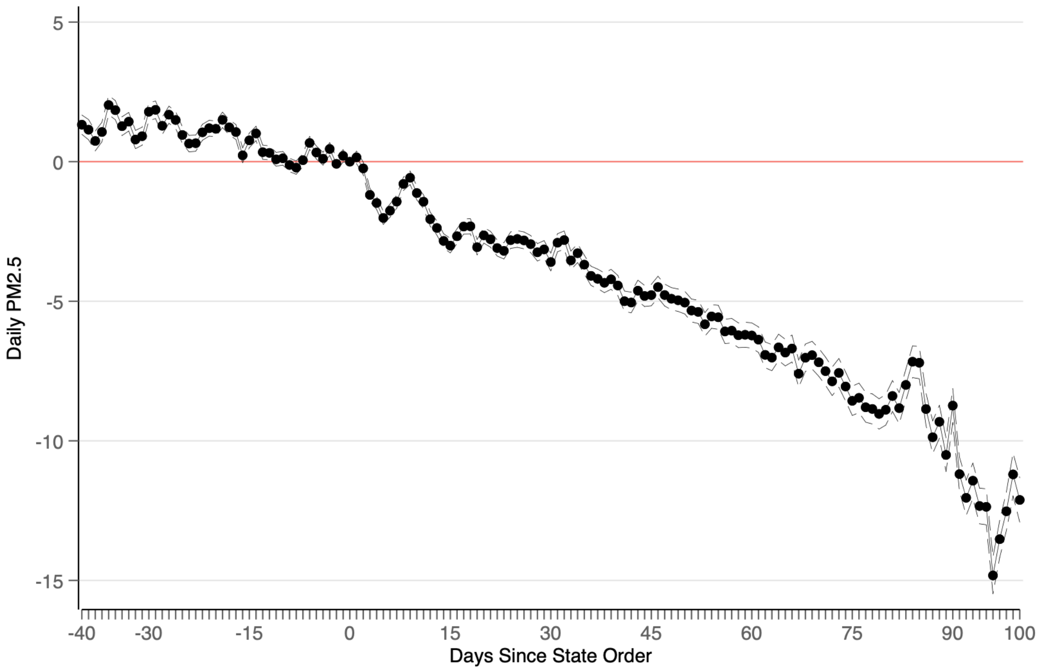
Notes: This figure presents regression coefficients for Air Quality Index (AQI) corresponding to number of days before/after state order issued. Date and county fixed effects included. We also include our full set of weather and policy controls. The lockdown date is presented with a red line. Confidence intervals at 95% presented.

FIGURE A8: PM2.5 Concentrations Over Time (Europe)



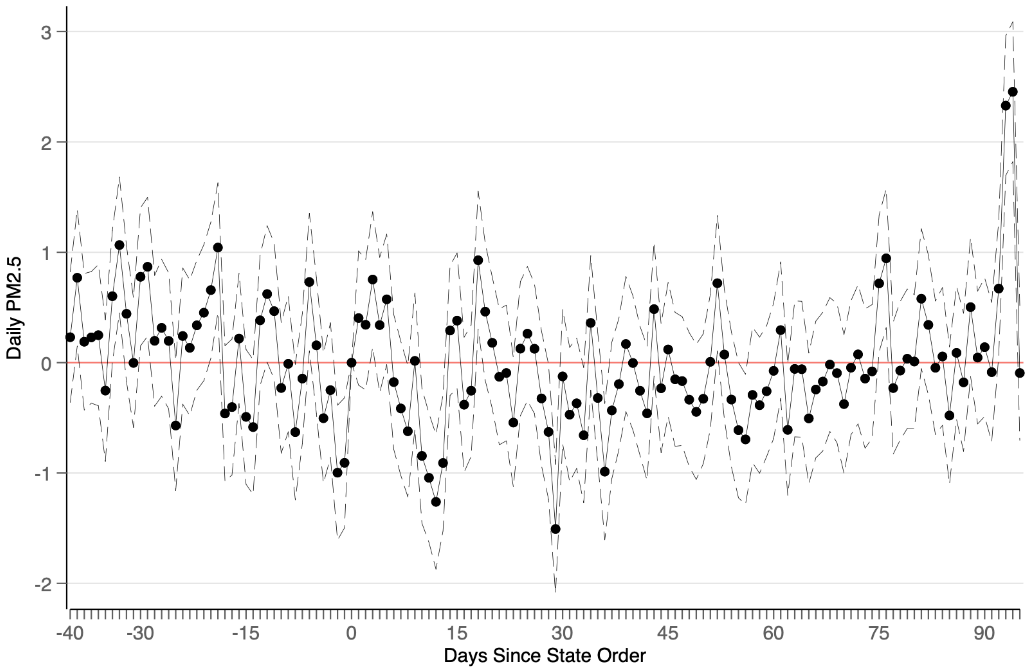
Notes: This figure presents regression coefficients for PM2.5 concentrations corresponding to number of days before/after national order issued. Area and date fixed effects included. We also include our full set of weather and policy controls. The baseline date is presented with a red line. Confidence intervals at 95% presented.

FIGURE A9: PM2.5 Concentrations Over Time - Synthetic Control



Notes: This figure presents the evolution of the difference between observed and the (simulated) counterfactual levels of PM2.5 expected from a synthetic control for each treated county, before and after state orders. Plotted are regression coefficients of days since state order issued. Significance bands at the 95% level. County fixed effects model. Date fixed effects also included.

FIGURE A10: PM2.5 Concentrations Over Time - Synthetic Placebo



Notes: The dependent variable is the difference between the observed PM2.5 levels of counties in states that did not issue a safer-at-home order in 2020 and those of a counterfactual county constructed from the remaining counties in the 8 untreated states. Plotted are regression coefficients of days since state order issued. Significance bands at the 95% level. County fixed effects model. Date fixed effects also included.

TABLE A1: State Orders and Pollution: State-Level

	(1) PM2.5	(2) PM2.5	(3) PM2.5	(4) PM2.5	(5) PM2.5
During safer-at-home-order	-1.476	-1.491	-1.486	-1.429	-1.152
	(0.439)	(0.440)	(0.438)	(0.458)	(0.371)
COVID cases per 10k		-0.005	-0.008	-0.005	0.006
		(0.004)	(0.012)	(0.015)	(0.015)
COVID deaths per 10k			0.039	-0.031	-0.153
			(0.135)	(0.165)	(0.164)
Constant	6.513	6.513	6.513	6.517	6.367
	(0.833)	(0.834)	(0.834)	(0.838)	(0.868)
State FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	5713	5713	5713	5713	5713
Counties	41	41	41	41	41

Notes: State orders significantly reduce PM2.5. The dependent variable is average daily PM2.5 concentration at the state-level. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors reported in parentheses. All columns include state and date fixed effects.

TABLE A2: State Orders and Pollution (PM2.5): Proactive and Reactive Counties

	All		Proactive		Reactive	
	(1)	(2)	(3)	(4)	(5)	(6)
During state order	-1.440	-1.372	-1.729	-1.170	-1.367	-1.286
	(0.372)	(0.413)	(0.455)	(0.609)	(0.398)	(0.420)
During county order			-0.501	-0.447		
			(0.287)	(0.302)		
COVID cases per 10k		-0.002		-0.013		-0.002
		(0.004)		(0.004)		(0.004)
COVID deaths per 10k		-0.006		-0.421		-0.008
		(0.035)		(0.427)		(0.034)
Constant	6.451	6.460	10.623	10.737	6.078	6.083
	(0.790)	(0.801)	(2.574)	(2.728)	(0.675)	(0.681)
Date FE	Y	Y	Y	Y	Y	Y
COVID-19 policies		Y		Y		Y
Weather controls	Y	Y	Y	Y	Y	Y
Observations	200094	200094	15220	15220	184874	184874
Counties	1592	1592	118	118	1474	1474

Notes: The dependent variable is average daily PM2.5 concentration at the county-level. Proactive counties implemented their own orders. Reactive counties had only state orders issued. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects. Sample restricted to counties within 50km of an air pollution monitoring station.

TABLE A3: State Orders and Pollution (PM2.5): Weighted Estimates

	(1) PM2.5	(2) PM2.5	(3) PM2.5	(4) PM2.5	(5) PM2.5
During safer-at-home-order	-2.522	-2.534	-2.549	-2.567	-2.069
	(0.681)	(0.663)	(0.660)	(0.638)	(0.453)
COVID cases per 10k		-0.008	-0.001	-0.001	0.001
		(0.004)	(0.007)	(0.008)	(0.008)
COVID deaths per 10k			-0.109	-0.126	-0.142
			(0.062)	(0.069)	(0.077)
Constant	11.200	11.199	11.199	11.215	10.394
	(2.013)	(2.024)	(2.024)	(2.027)	(1.980)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	200094	200094	200094	200094	200094
Counties	1592	1592	1592	1592	1592

Notes: State orders significantly reduce PM2.5, weighted by county population. The dependent variable is average daily PM2.5 concentration at the county-level. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 100 people. Standard errors clustered at the state-level reported in parentheses, clustered at the state-level. All columns include county and date fixed effects.

TABLE A4: State Orders and Air Quality Index

	(1) AQI	(2) AQI	(3) AQI	(4) AQI	(5) AQI
During safer-at-home-order	-6.240	-6.251	-6.251	-6.003	-5.093
	(0.307)	(0.308)	(0.308)	(0.326)	(0.288)
COVID cases per 10k		-0.012	-0.012	-0.013	-0.008
		(0.005)	(0.008)	(0.008)	(0.008)
COVID deaths per 10k			-0.000	-0.022	-0.034
			(0.100)	(0.107)	(0.105)
Constant	25.591	25.591	25.591	25.611	24.695
	(0.436)	(0.436)	(0.436)	(0.438)	(0.438)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	200094	200094	200094	200094	200094
Counties	1592	1592	1592	1592	1592

Notes: State orders significantly reduce Air Quality Index. The dependent variable is average daily AQI. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects.

TABLE A5: State Orders and Air Quality Risk

	(1) AQI>50	(2) AQI>50	(3) AQI>50	(4) AQI>50	(5) AQI>50
During safer-at-home-order	-0.098	-0.098	-0.098	-0.093	-0.084
	(0.031)	(0.031)	(0.031)	(0.033)	(0.030)
COVID cases per 10k		-0.000	-0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
COVID deaths per 10k			-0.001	-0.002	-0.002
			(0.002)	(0.002)	(0.002)
Constant	0.097	0.097	0.097	0.097	0.082
	(0.036)	(0.036)	(0.036)	(0.037)	(0.036)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	200094	200094	200094	200094	200094
Counties	1592	1592	1592	1592	1592

Notes: State orders significantly reduce Air Quality Risk of being in 'code yellow' or above. The dependent variable indicates if average daily AQI exceeds 50, the threshold for moderate health effects from air pollution. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects.

TABLE A6: State Orders and Aerosol Optical Depth (Restricted)

	(1) AOD	(2) AOD	(3) AOD	(4) AOD	(5) AOD
During safer-at-home-order	-2.065	-2.055	-2.165	-1.510	1.092
	(6.909)	(6.904)	(6.905)	(6.365)	(5.233)
COVID cases per 10k		0.019	0.094	0.082	0.091
		(0.043)	(0.070)	(0.070)	(0.066)
COVID deaths per 10k			-1.448	-1.717	-1.775
			(0.946)	(0.964)	(0.938)
Constant	38.944	38.937	38.965	38.897	18.538
	(5.181)	(5.171)	(5.170)	(5.158)	(6.173)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	41697	41697	41697	41697	41697
Counties	1541	1541	1541	1541	1541

Notes: State orders reduce aerosol optical depth. The dependent variable is aerosol optical depth. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects. The sample is restricted to counties within 50km of a monitoring station.

TABLE A7: State Orders and Aerosol Optical Depth (All)

	(1) AOD	(2) AOD	(3) AOD	(4) AOD	(5) AOD
During safer-at-home-order	-2.832	-2.823	-2.852	-2.328	-0.541
	(5.276)	(5.273)	(5.277)	(5.103)	(4.264)
COVID cases per 10k		0.025	0.056	0.051	0.054
		(0.033)	(0.052)	(0.050)	(0.047)
COVID deaths per 10k			-0.717	-0.873	-0.804
			(0.778)	(0.765)	(0.745)
Constant	35.593	35.585	35.596	35.473	16.872
	(5.485)	(5.479)	(5.478)	(5.492)	(7.376)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	77279	77279	77279	77279	77279
Counties	2532	2532	2532	2532	2532

Notes: State orders significantly reduce aerosol optical depth. The dependent variable is aerosol optical depth. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects.

TABLE A8: State Orders and Pollution (PM2.5) Alternative Estimators

	(1) Baseline	(2) Simple Average	(3) Dynamic Average	(4) Group Average
During safer-at-home-order	-1.671	-1.027	-0.749	-0.449
	(0.084)	(0.157)	(0.164)	(0.144)
County FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Observations	200094	200094	200094	200094
Counties	1592	1592	1592	1592

Notes: State orders significantly reduce PM2.5. The dependent variable is average daily PM2.5 concentration at the county-level. Column 1 contains the baseline estimate. Columns 2–4 provide alternative estimators based on Callaway and Sant'Anna, Forthcoming: the "simple" aggregated average treatment on the treated (ATT), the "dynamic" aggregated ATT, and the "group" aggregated ATT, respectively. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. Robust standard errors clustered at the state-level reported in parentheses. All columns include county and date fixed effects. Sample restricted to counties within 50km of an air pollution monitoring station.

TABLE A9: Stringency of State Orders

	PM2.5		Collisions	
	(1)	(2)	(3)	(4)
Less Stringent Order	-0.194	-0.110	0.899	0.872
	(0.714)	(0.592)	(0.008)	(0.008)
More Stringent Order	-0.903	-0.778	0.937	0.897
	(0.656)	(0.539)	(0.009)	(0.008)
COVID-19 Cases	-0.003	-0.002	1.002	1.002
	(0.004)	(0.004)	(0.000)	(0.000)
COVID-19 Deaths	0.002	-0.003	0.985	0.987
	(0.034)	(0.034)	(0.002)	(0.002)
Constant	6.693	6.462		
	(0.163)	(0.163)		
County FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
COVID-19 policies	Y	Y	Y	Y
Weather controls		Y		Y
Observations	190370	190370	270875	270714
Counties	1592	1592	1894	1893

Notes: The dependent variable is average daily PM2.5 concentration at the county-level for columns 1 and 2 and count of traffic collisions at the county-level for columns 3 and 4, respectively. The time period spans January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses for columns 1 and 2. Robust standard errors reported in parentheses for columns 3 and 4. All columns include county and date fixed effects.

TABLE A10: State Orders and Pollution Heterogeneity (PM2.5)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Urban	Rural	Majority Trump	Minority Trump	Above Median 65+	Below Median 65+	Above Median Work from Home	Below Median Work from Home
During order	-1.484	-1.200	-1.159	-1.762	-0.945	-1.629	-1.403	-1.282
	(0.463)	(0.371)	(0.404)	(0.408)	(0.451)	(0.433)	(0.572)	(0.485)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
COVID-19 policies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	109333	90761	142946	57148	75769	124325	110433	89661
Counties	816	776	1169	423	637	955	849	743

Notes: The dependent variable is average daily PM2.5 concentration at the county-level. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors clustered at the state-level reported in parentheses. All columns include date fixed effects. Sample restricted to counties within 50km of an air pollution monitoring station. Columns 1–2 are the effects of the order for urban and rural counties, respectively. Columns 2–4 are for counties in which a majority voters voted for President Trump. Columns 5–6 are for counties with with an above and below median share of population aged 65 and over, respectively. Columns 7–8 are for counties in states with above and below median shares of occupations that can be done remotely, respectively.

TABLE A11: Characteristics of Trump and non-Trump Majority Counties (Pre-Order Time Period)

Variable	(1) Non-Trump N	Mean/SD	(2) Trump N	Mean/SD	T-test Difference (1)-(2)
PM2.5	23063	6.649	131957	6.398	0.250
		(3.602)		(3.679)	
PM2.5 > 12	23063	0.082	131957	0.065	0.016
		(0.274)		(0.247)	
Travel distance	22669	-0.118	127003	-0.106	-0.011
		(0.195)		(0.180)	
Urban County	23101	0.700	132433	0.307	0.392
		(0.458)		(0.461)	
Above median population 65+	23101	0.254	132433	0.582	-0.328
		(0.435)		(0.493)	
Above median work from home	23101	0.564	132433	0.410	0.154
		(0.496)		(0.492)	
COVID cases per 10k	23101	5.759	132433	6.829	-1.070
		(30.020)		(37.699)	
COVID deaths per 10k	23101	0.087	132433	0.131	-0.044
		(0.435)		(0.784)	

Notes: The value displayed for t-tests are the differences in the means across the groups.

TABLE A12: European National Orders and Pollution (PM2.5) By Country

	(1) FR	(2) GE	(3) IT	(4) SP	(5) UK
During lockdown	-14.649	-44.275	-4.128	-15.736	-21.861
	(0.592)	(0.221)	(0.572)	(0.367)	(0.379)
Constant	22.406	48.193	15.604	17.531	27.071
	(0.416)	(0.094)	(0.279)	(0.263)	(0.159)
Date FE	Y	Y	Y	Y	Y
Temperature	Y	Y	Y	Y	Y
Precipitation	Y	Y	Y	Y	Y
Observations	23880	66276	28432	9010	25188

Notes: National orders significantly reduce PM2.5 in France, Germany, Italy, Spain and the United Kingdom. The dependent variable is average daily PM2.5 concentration, measured at the population centroid for an administrative area from the three nearest air quality monitors. The mean concentration during the period is 8.7, and the within-unit standard deviation is 6.0. An observation is an area-day. A total of 841 areas are used with 96 from France (departments), 403 from Germany (kreise), 113 from Italy (provinces), 55 from Spain (provinces) and 192 from the United Kingdom (counties). Observations are from the first to the 300th day of 2020, which includes both before, during, and after lockdowns. Standard errors reported in parentheses. All columns include area and date fixed effects.

TABLE A13: State Orders and Collisions: State-Level

	(1)	(2)	(3)	(4)	(5)
During safer-at-home-order	0.840	0.839	0.834	0.823	0.797
	(0.109)	(0.114)	(0.110)	(0.095)	(0.092)
COVID cases per 10k		1.003	1.010	1.009	1.008
		(0.001)	(0.004)	(0.003)	(0.003)
COVID deaths per 10k			0.908	0.919	0.930
			(0.034)	(0.029)	(0.029)
State FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	5716	5716	5716	5716	5716
Counties	41	41	41	41	41

Notes: State orders significantly reduce traffic collisions. Poisson model with fixed effects. The dependent variable is count of traffic collisions at the state-level. Coefficients are incidence rate-ratios, wherein a value below one indicates a decrease in the dependent variable and a value above indicates an increase. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors reported in parentheses. All columns include state and date fixed effects.

TABLE A14: State Orders and Collisions: Proactive and Reactive Counties

	All		Proactive		Reactive	
	(1)	(2)	(3)	(4)	(5)	(6)
During order	0.806	0.793	0.752	0.750	0.865	0.850
	(0.049)	(0.045)	(0.086)	(0.067)	(0.053)	(0.054)
During county order			0.762	0.756		
			(0.057)	(0.048)		
COVID cases per 10k		1.002		1.007		1.002
		(0.001)		(0.005)		(0.001)
COVID deaths per 10k		0.983		0.809		0.979
		(0.011)		(0.098)		(0.010)
County FE	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y
COVID-19 policies		Y		Y		Y
Weather controls	Y	Y	Y	Y	Y	Y
Observations	237408	237408	16479	16479	220929	220929
Counties	1710	1710	121	121	1589	1589

Notes: The dependent variable is average daily collisions at the county-level. Proactive counties implemented their own orders. Reactive counties had only state orders issued. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors reported in parentheses. All columns include county and date fixed effects.

TABLE A15: State Orders and Collision Heterogeneity

	(1) Urban	(2) Rural	(3) Majority Trump	(4) Minority Trump	(5) Above Median 65+	(6) Below Median 65+	(7) Above Median Work from Home	(8) Below Median Work from Home
During order	0.774	1.077	0.881	0.745	0.776	0.804	0.706	1.085
	(0.046)	(0.099)	(0.051)	(0.053)	(0.060)	(0.047)	(0.049)	(0.102)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
COVID-19 policies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	124770	112638	171226	66182	98950	138458	142019	95389
Counties	889	821	1247	463	704	1006	1004	706

Notes: The dependent variable is count of traffic collisions at the county-level. Coefficients are incidence rate-ratios, wherein a value below one indicates a decrease in the dependent variable and a value above indicates an increase. The time period is January 1st, 2020 through the moment the statewide safer-at-home ended. COVID-19 known cases and deaths per 10,000 people. Robust standard errors reported in parentheses. All columns include county and date fixed effects. Columns 1–2 are the effects of the order for urban and rural counties, respectively. Columns 3–4 are for counties in which a majority voters voted for President Trump. Columns 5–6 are for counties with with an above and below median share of population aged 65 and over, respectively. Columns 7–8 are for counties in states with above and below median shares of occupations that can be done remotely, respectively.

TABLE A16: State Orders and Pollution (PM2.5): Simulation Using Synthetic Control

	(1) synth_diff	(2) synth_diff	(3) synth_diff	(4) synth_diff	(5) synth_diff
During order	-1.060	-1.057	-1.059	-1.072	-0.884
	(0.397)	(0.397)	(0.397)	(0.419)	(0.339)
COVID cases per 10k		0.002	0.002	0.002	0.002
		(0.001)	(0.001)	(0.001)	(0.001)
COVID deaths per 10k			-0.009	-0.002	0.003
			(0.030)	(0.027)	(0.031)
Constant	1.045	1.044	1.044	1.044	0.757
	(0.866)	(0.866)	(0.866)	(0.869)	(0.851)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	237243	237243	237243	237243	237243
Counties	1418	1418	1418	1418	1418

Notes: State orders significantly reduce PM2.5. The dependent variable is the *difference* between observed and counterfactual PM2.5 at the county-day level. For each county that issued a lockdown order, a synthetic counterfactual county was created from a weighted convex combination of counties from the 8 'donor' states that did not issue an order. Synthetic weights determined by pre-treatment period PM2.5 levels. The time period is January 1st, 2020 through June 30th, 2020. COVID-19 known cases and deaths per 10,000 people using 2019 population estimates. Standard errors reported in parentheses, clustered at the state-level. All columns include state and date fixed effects. Sample restricted to counties within 50km of an air pollution monitoring station.

TABLE A17: State Orders and Pollution (PM2.5): Simulation Using Synthetic Control Placebo

	(1)	(2)	(3)	(4)	(5)
During order (placebo)	-0.705	-0.694	-0.694	-0.720	1.084
	(0.962)	(0.966)	(0.965)	(0.978)	(2.237)
COVID cases per 10k		-0.002	-0.003	-0.003	-0.003
		(0.002)	(0.002)	(0.002)	(0.002)
COVID deaths per 10k			0.052	0.055	0.066
			(0.066)	(0.065)	(0.064)
Constant	0.301	0.301	0.300	0.300	0.369
	(0.942)	(0.945)	(0.944)	(0.934)	(0.940)
County FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
COVID-19 policies				Y	Y
Weather controls					Y
Observations	30989	30989	30989	30989	30989
Counties	209	209	209	209	209

Notes: Placebo state orders do not significantly affect PM2.5. The dependent variable is the *difference* between observed and counterfactual PM2.5 at the county-day level. For each county in each of the 8 states that did not issue a lockdown order, a synthetic counterfactual county was created from a weighted convex combination of remaining counties in the 8 'donor' states that did not issue an order. Synthetic weights determined by pre-treatment period PM2.5 levels. The time period is January 1st, 2020 through June 30th, 2020. COVID-19 known cases and deaths per 10,000 people using 2019 population estimates. Standard errors reported in parentheses, clustered at the state-level. All columns include state and date fixed effects. Sample restricted to counties within 50km of an air pollution monitoring station.

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program”. In: *Journal of the American statistical Association* 105.490, pp. 493–505.

— (2011). “Synth: An r Package for Synthetic Control Methods in Comparative Case Studies”. In: *Journal of Statistical Software* 42.13.

Abadie, Alberto and Javier Gardeazabal (2003). “The Economic Costs of Conflict: A Case Study of the Basque Country”. In: *American Economic Review* 93.1, pp. 113–132.

Adams-Prassl, Abi et al. (2020). *The Impact of the Coronavirus Lockdown on Mental Health: Evidence from the US*. Working Papers 2020-030, Human Capital and Economic Opportunity Working Group.

Akinnibi, Fola, Sarah Holder, and Christopher Cannon (2021). *Cities Say They Want to Defund the Police. Their Budgets Say Otherwise.* URL: https://www.bloomberg.com/graphics/2021-city-budget-police-funding/.

Alesina, Alberto and Eliana La Ferrara (2014). “A Test of Racial Bias in Capital Sentencing”. In: *American Economic Review* 104.11, pp. 3397–3433.

Anderson, Elijah (1999). *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. WW Norton & Company.

Andrews, Rafe H. et al. (2019). *’When Deported, You Become Nothing’*. URL: https://www.politico.com/magazine/story/2019/04/19/displaced-puebla-deportation-immigration-new-york-photos-226657.

Ang, Desmond et al. (2021). “Police violence reduces civilian cooperation and engagement with law enforcement”. In: *HKS Working Paper No. RWP21-022*.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson (2012). “The Impact of Jury Race in Criminal Trials”. In: *Quarterly Journal of Economics* 127.2, pp. 1017–1055.

Avdan, Nazli (2014). “Do Asylum Recognition Rates in Europe Respond to Transnational Terrorism? The Migration-Security Nexus Revisited”. In: *European Union Politics* 15.4, pp. 445–471.

Ba, Bocar A et al. (2021). “The role of officer race and gender in police-civilian interactions in Chicago”. In: *Science* 371.6530, pp. 696–702.

Barwick, Panle Jia et al. (2017). *Air Pollution, Health Spending and Willingness to Pay for Clean Air in China*. SSRN Electronic Journal.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins (2009). “Migration and Hedonic Valuation: The Case of Air Quality”. In: *Journal of Environmental Economics and Management* 58.1, pp. 1–14.

Bayer, Patrick et al. (2016). “A Dynamic Model of Demand for Houses and Neighborhoods”. In: *Econometrica* 84.3, pp. 893–942.

Becker, Gary S (1968). “Crime and Punishment: An Economic Approach”. In: *Journal of Political Economy* 76.2, pp. 169–217.

Becker, Gary S. and Yona Rubinstein (2011). *Fear and the Response to Terrorism: An Economic Analysis*. CEP Discussion Paper 1079.

Beland, Louis-Philippe and Daniel A Brent (2018). "Traffic and Crime". In: *Journal of Public Economics* 160, pp. 96–116.
Beland, Louis-Philippe, Abel Brodeur, and Taylor Wright (2020). *COVID-19, Stay-at-Home Orders and Employment: Evidence from CPS Data*. GLO Working Paper 559.
Berg, Mark T et al. (2016). "Cynical streets: Neighborhood social processes and perceptions of criminal injustice". In: *Criminology* 54.3, pp. 520–547.
Bisin, Alberto et al. (2008). "Are Muslim Immigrants Different in Terms of Cultural Integration?" In: *Journal of the European Economic Association* 6.2-3, pp. 445–456.
Blomberg, S Brock, Gregory D Hess, and Athanasios Orphanides (2004). "The Macroeconomic Consequences of Terrorism". In: *Journal of Monetary Economics* 51.5, pp. 1007–1032.
Bobo, Lawrence D and Victor Thompson (2006). "Unfair by design: The war on drugs, race, and the legitimacy of the criminal justice system". In: *Social Research: An International Quarterly* 73.2, pp. 445–472.
Bowe, Benjamin et al. (2019). "Burden of Cause-Specific Mortality Associated With PM2. 5 Air Pollution in the United States". In: *JAMA Network Open* 2.11, e1915834–e1915834.
Brodeur, A et al. (2021). "COVID-19, Lockdowns and Well-Being: Evidence from Google Trends". In: *Journal of Public Economics* 193.
Brodeur, Abel (2018). "The Effect of Terrorism on Employment and Consumer Sentiment: Evidence from Successful and Failed Terror Attacks". In: *American Economic Journal: Applied Economics* 10.4, pp. 246–82.
Brodeur, Abel and Marie Connolly (2013). "Do Higher Child Care Subsidies Improve Parental Well-Being? Evidence from Quebec's Family Policies". In: *Journal of Economic Behavior & Organization* 93, pp. 1–16.
Brodeur, Abel, Nikolai Cook, and Taylor Wright (2020). *On the Effects of COVID-19 Safer-At-Home Policies on Social Distancing, Car Crashes and Pollution*. IZA Discussion Paper 13255.
Brodeur, Abel, Idaliya Grigoryeva, and Lamis Kattan (2020). *Stay-At-Home Orders, Social Distancing and Trust*. IZA Discussion Paper 13234.
Brodeur, Abel et al. (2020). *A Literature Review of the Economics of COVID-19*. IZA Discussion Paper 13411.
Brodzinsky, Sibylla (2015). *The migrants who fled violence for the US only to be sent back to their deaths*. URL: https://www.theguardian.com/world/2015/oct/12/deportation-migrants-flee-honduras-guatemala-salvador.
Brunson, Rod K (2007). ""Police don't like black people": African-American young men's accumulated police experiences". In: *Criminology & Public Policy* 6.1, pp. 71–101.
Bullinger, Lindsey et al. (2020). *The Neglected Ones: Time at Home During COVID-19 and Child Maltreatment*. SSRN 3674064.
Bulman, May (2018). *Afghan father who sought refuge in UK 'shot dead by Taliban' after being deported by Home Office*. URL: https://www.independent.co.uk/news/uk/home-news/zainadin-fazlie-deport-home-office-taliban-afghanistan-shot-dead-refugee-a8536736.html.
Cable News Network (CNN) (2006). *Courts Told to Be Nicer to Immigrants*. Online; accessed March 2018.
Callaway, Brantly and Pedro H.C. Sant'Anna (2020). *DID: Difference in Differences*. R package version 2.0.0. URL: https://bcallaway11.github.io/did/.
Callaway, Brantly and Pedro HC Sant'Anna (Forthcoming). "Difference-in-Differences with Multiple Time Periods". In: *Journal of Econometrics*.

Campbell, Bradley A, Justin Nix, and Edward R Maguire (2018). "Is the number of citizens fatally shot by police increasing in the post-Ferguson era?" In: *Crime & Delinquency* 64.3, pp. 398–420.

Card, David and Gordon B Dahl (2011). "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior". In: *Quarterly Journal of Economics* 126.1, pp. 103–143.

Carr, Jillian B and Jennifer L Doleac (2018). "Keep the Kids Inside? Juvenile Curfews and Urban Gun Violence". In: *Review of Economics and Statistics* 100.4, pp. 609–618.

Carr, Patrick J, Laura Napolitano, and Jessica Keating (2007). "We never call the cops and here is why: A qualitative examination of legal cynicism in three Philadelphia neighborhoods". In: *Criminology* 45.2, pp. 445–480.

Chattopadhyay, Sudip (1999). "Estimating the Demand for Air Quality: New Evidence Based on the Chicago Housing Market". In: *Land Economics*, pp. 22–38.

Chay, Kenneth Y and Michael Greenstone (2005). "Does Air Quality Matter? Evidence from the Housing Market". In: *Journal of Political Economy* 113.2, pp. 376–424.

Cicala, Steve et al. (2020). *Expected Health Effects of Reduced Air Pollution from COVID-19 Social Distancing*. NBER Working Paper 27135.

Cohen, Alma and Crystal S Yang (2019). "Judicial Politics and Sentencing Decisions". In: *American Economic Jounral: Economic Policy* 11.1, pp. 160–91.

Cornelissen, Thomas and Uwe Jirjahn (2012). "September 11th and the Earnings of Muslims in Germany–The Moderating Role of Education and Firm Size". In: *Journal of Economic Behavior & Organization* 81.2, pp. 490–504.

Count Love (2021). *Count Love*. URL: https://countlove.org/.

CPD (2017). *General Order G03-02 Use of Force*. Tech. rep. Chicago Police Department.

Crain, Nicole V and W Mark Crain (2006). "Terrorized Economies". In: *Public Choice* 128.1-2, pp. 317–349.

Currie, Janet et al. (2015). "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings". In: *American Economic Review* 105.2, pp. 678–709.

Dang, Hai-Anh H and Trong-Anh Trinh (Forthcoming). "Does the COVID-19 Lockdown Improve Global Air Quality? New Cross-National Evidence on its Unintended Consequences". In: *Journal of Environmental Economics and Management*.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso (2011). "Extraneous Factors in Judicial Decisions". In: *Proceedings of the National Academy of Sciences* 108.17, pp. 6889–6892.

Dávila, Alberto and Marie T Mora (2005). "Changes in the Earnings of Arab Men in the US Between 2000 and 2002". In: *Journal of Population Economics* 18.4, pp. 587–601.

De Chaisemartin, Clement and Xavier d'Haultfoeuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". In: *American Economic Review* 110.9, pp. 2964–96.

Department of Justice (2018a). *About the Office*. Executive Office for Immigration Review. Online; accessed August 2018.

— (2018b). *Immigration Judge*. Executive Office for Immigration Review. Online; accessed March 2018.

Deschênes, Olivier, Michael Greenstone, and Joseph S Shapiro (2017). "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program". In: *American Economic Review* 107.10, pp. 2958–89.

DeSilver, Drew, Michael Lipka, and Dalia Fahmy (2020). "10 things we know about race and policing in the US". In: *Pew Research, June* 3, p. 2020.

Desmond, Matthew, Andrew V Papachristos, and David S Kirk (2016). "Police violence and citizen crime reporting in the black community". In: *American Sociological Review* 81.5, pp. 857–876.

— (2020). "Evidence of the effect of police violence on citizen crime reporting". In: *American Sociological Review* 85.1, pp. 184–190.

Dingel, Jonathan I and Brent Neiman (2020). *How Many Jobs Can be Done at Home?* NBER Working Paper 26948.

Dreher, Axel, Martin Gassebner, and Paul Schaudt (2017). *The Effect of Migration on Terror-Made at Home or Imported from Abroad?* CEPR Discussion Paper 12062.

Elsayed, Ahmed and Andries De Grip (2018). "Terrorism and the Integration of Muslim Immigrants". In: *Journal of Population Economics* 31.1, pp. 45–67.

Engle, Sam, John Stromme, and Anson Zhou (2020). "Staying at Home: Mobility Effects of Covid-19". In: *CEPR Covid Economics* 4, pp. 86–102.

Eren, Ozkan and Naci Mocan (2018). "Emotional Judges and Unlucky Juveniles". In: *American Economic Journal: Applied Economics* 10.3, pp. 171–205.

Gaston, Shytierra, Jamein P Cunningham, and Rob Gillezeau (2019). "A Ferguson effect, the drug epidemic, both, or neither? Explaining the 2015 and 2016 US homicide rises by race and ethnicity". In: *Homicide Studies* 23.3, pp. 285–313.

Gau, Jacinta M and Rod K Brunson (2010). "Procedural justice and order maintenance policing: A study of inner-city young men's perceptions of police legitimacy". In: *Justice Quarterly* 27.2, pp. 255–279.

Goodman-Bacon, Andrew and Jan Marcus (2020). *Difference-in-Differences to Identify Causal Effects of COVID-19 Policies*. DIW Berlin Working Paper.

Gould, Eric D and Esteban F Klor (2016). "The Long-Run Effect of 9/11: Terrorism, Backlash, and the Assimilation of Muslim Immigrants in the West". In: *Economic Journal* 126.597, pp. 2064–2114.

Graf, Christoph, Federico Quaglia, and F A Wolak (Forthcoming). "(Machine) Learning from the COVID-19 Lockdown about Electricity Market Performance with a Large Share of Renewables". In: *Journal of Environmental Economics and Management*.

Gwynn, David W (1967). "Relationship of Accident Rates and Accident Involvements with Hourly Volumes". In: *Traffic Quarterly* 21.3.

Hall, Ola et al. (2019). "Population Centroids of the World Administrative Units from Nighttime Lights 1992-2013". In: *Scientific Data* 6.1, pp. 1–8.

Han, Seunghee, Jennifer S. Lerner, and Dacher Keltner (2007). "Feelings and Consumer Decision Making: The Appraisal-Tendency Framework". In: *Journal of Consumer Psychology* 17.3, pp. 158–168. ISSN: 10577408.

Harrison, David and Daniel L Rubinfeld (1978). "The Air Pollution and Property Value Debate: Some Empirical Evidence". In: *Review of Economics and Statistics*, pp. 635–638.

He, Guojun, Yuhang Pan, and Takanao Tanaka (2020). "The Short-Term Impacts of COVID-19 Lockdown on Urban Air Pollution in China". In: *Nature Sustainability*, pp. 1–7.

Head, J Al (1959). "Predicting Traffic Accidents from Roadway Elements on Urban Extensions of State Highways". In: *Highway Research Board Bulletin* 208, pp. 45–63.

Heyes, Anthony and Soodeh Saberian (2019). "Temperature and Decisions: Evidence from 207,000 Court Cases". In: *American Economic Journal: Applied Economics* 11.2, pp. 238–65.

Hoekstra, Mark and CarlyWill Sloan (2020). "Does race matter for police use of force? Evidence from 911 calls". In: *NBER Working Paper No. 26774*.

Hsiang, Solomon et al. (2020). *The Effect of Large-Scale Anti-Contagion Policies on the Coronavirus (COVID-19) Pandemic*. medRxiv.

Ito, Koichiro and Shuang Zhang (2020). "Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China". In: *Journal of Political Economy* 128.5, pp. 000–000.

Jayachandran, Seema and Adriana Lleras-Muney (2009). "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines". In: *Quarterly Journal of Economics* 124.1, pp. 349–397.

Johnson, Eric J and Amos Tversky (1983). "Affect, Generalization, and the Perception of Risk". In: *Journal of Personality and Social Psychology* 45.1, pp. 20–31.

Kaminski, Robert J. and Eric S. Jefferis (1998). "The Effect of a Violent Televised Arrest on Public Perceptions of the Police". In: *Policing: An International Journal of Police Strategies & Management* 21.4, pp. 683–706. DOI: 10.1108/13639519810241692.

Kaushal, Neeraj, Robert Kaestner, and Cordelia Reimers (2007). "Labor Market Effects of September 11th on Arab and Muslim Residents of the United States". In: *Journal of Human Resources* 42.2, pp. 275–308.

Kirk, David S and Andrew V Papachristos (2011). "Cultural mechanisms and the persistence of neighborhood violence". In: *American Journal of Sociology* 116.4, pp. 1190–1233.

Kochel, Tammy (2015a). "Assessing the initial impact of the Michael Brown shooting and police and public responses to it on St Louis county residents' views about police". In: *Carbondale:Southern Illinois University Carbondale*.

— (2015b). "Ferguson's long term impact on public views about police". In: *Carbondale: Southern Illinois University Carbondale*.

Kong, Edward and Daniel Prinz (2020). "Disentangling Policy Effects Using Proxy Data: Which Shutdown Policies Affected Unemployment during the COVID-19 Pandemic?" In: *Journal of Public Economics* 189, p. 104257.

Kwak, Hyounggon, Rick Dierenfeldt, and Susan McNeeley (2019). "The code of the street and cooperation with the police: Do codes of violence, procedural injustice, and police ineffectiveness discourage reporting violent victimization to the police?" In: *Journal of Criminal Justice* 60, pp. 25–34.

Lane, Emily (2016). *Why do police shoot so many times? FBI, experts answer on officer-involved shootings*. URL: https://www.nola.com/news/crime_police/article_ae82835c-0212-5e50-a175-85601a1ed8bb.html.

Le Quéré, Corinne et al. (2020). "Temporary Reduction in Daily Global CO 2 Emissions during the COVID-19 Forced Confinement". In: *Nature Climate Change*, pp. 1–7.

Leach, Andrew, Nic Rivers, and Blake Shaffer (2020). "Canadian Electricity Markets During the COVID-19 Pandemic: An Initial Assessment". In: *Canadian Public Policy* 46.S2, S145–S159.

Lerner, Jennifer S. et al. (2015). "Emotion and Decision Making". In: *Annual Review of Psychology* 66.1, pp. 799–823.

Lerner, J.S. et al. (2003). "Effects of Fear and Anger on Perceived Risks of Terrorism: A National Field Experiment". In: *Psychological Science* 14.2, pp. 144–150. ISSN: 09567976.

Leslie, Emily and Riley Wilson (2020). "Sheltering in Place and Domestic Violence: Evidence from Calls for Service During COVID-19". In: *Journal of Public Economics* 189.

Liptak, Adam (2005). *Courts Criticize Judges' Handling of Asylum Cases*. The New York Times. Online; accessed March 2018.

Mapping Police Violence (2021). *Mapping Volice Violence*. URL: https://mappingpoliceviolence.org/.

Mason, Patrick L and Andrew Matella (2014). "Stigmatization and Racial Selection After September 11, 2001: Self-Identity Among Arab and Islamic Americans". In: *IZA Journal of Migration* 3.1, p. 20.

McCarthy, Bill, John Hagan, and Daniel Herda (2020). "Neighborhood climates of legal cynicism and complaints about abuse of police power". In: *Criminology* 58.3, pp. 510–536.

McConnell, Brendan and Imran Rasul (2019). *Contagious Animosity in the Field: Evidence from the Federal Criminal Justice System*. Working Paper.

Miller, Banks, Linda Camp Keith, and Jennifer S Holmes (2015). *Immigration Judges and U.S. Asylum Policy*. University of Pennsylvania Press.

Mohammad, Saif M and Peter D Turney (2013). "Crowdsourcing a word–emotion association lexicon". In: *Computational Intelligence* 29.3, pp. 436–465.

Moosavi, Sobhan et al. (2019). *A Countrywide Traffic Accident Dataset*. arXiv, eprint 1906.05409. eprint: 1906.05409 (cs.DB).

National Association of Immigration Judges (2011). *Statement of National Association of Immigration Judges Before the Senate Committee on the Judiciary on "Improving Efficiency and Ensuring Justice in the Immigration Court System"*.

Oglesby-Neal, Ashlin, Emily Tiry, and K Kim (2019). "Public perceptions of police on social media". In: *Urban Institute*, pp. 1–12.

Ortiz, Aimee (2020). "Confidence in police is at record low, Gallup survey finds". In: *Sec. US The New York Times. August* 12.

Pansini, Riccardo and Davide Fornacca (2020). *Higher Virulence of COVID-19 in the Air-Polluted Regions of Eight Severely Affected Countries*. medRxiv.

Park, Madison, Emanuella Grinberg, and Holly Yan (2017). *Minneapolis woman killed by police: What we know*. URL: https://www.cnn.com/2017/07/18/us/minneapolis-woman-shooting-what-we-know/index.html.

Philippe, Arnaud and Aurélie Ouss (2018). ""No Hatred or Malice, Fear or Affection": Media and Sentencing". In: *Journal of Political Economy* 126.5, pp. 2134–2178.

Raifman, Julia et al. (2020). *COVID-19 US State Policy Database*. MI: Inter-university Consortium for Political and Social Research.

Ramji-Nogales, Jaya, Andrew I. Schoenholtz, and Philip G. Schrag (2007). "Refugee Roulette: Disparities in Asylum Adjudication". In: *Stanford Law Review* 60.2, pp. 295–411.

Ratcliffe, Anita and Stephanie von Hinke Kessler Scholder (2015). "The London Bombings and Racial Prejudice: Evidence from the Housing and Labor Market". In: *Economic Inquiry* 53.1, pp. 276–293.

Refugee, Asylum, and International Operations Directorate (2016). *Affirmative Asylum Procedures Manual*. Tech. rep.

Retallack, Angus Eugene and Bertram Ostendorf (2019). "Current Understanding of the Effects of Congestion on Traffic Accidents". In: *International Journal of Environmental Research and Public Health* 16.18, p. 3400.

Sanga, Sarath and Justin McCrary (2020). *The Impact of the Coronavirus Lockdown on Domestic Violence*. SSRN 3612491.

Schoppert, David W (1957). "Predicting Traffic Accidents from Roadway Elements of Rural Two-Lane Highways with Gravel Shoulders". In: *Highway Research Board Bulletin* 158, pp. 4–26.

Shayo, Moses and Asaf Zussman (2011). "Judicial Ingroup Bias in the Shadow of Terrorism". In: *Quarterly Journal of Economics* 126.3, pp. 1447–1484.

Stillman, Sarah (2018). *When Deportation Is a Death Sentence*. URL: {https://www.newyorker.com/magazine/2018/01/15/when-deportation-is-a-death-sentence}.

Sunshine, Jason and Tom R. Tyler (2003). "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing". In: *Law & Society Review* 37.3, pp. 513–548. DOI: 10.1111/1540-5893.3703002.

TASER (2009). *Training Bulletin 15.0 Medical Research Update and Revised Warnings*. URL: https://media.cdn.lexipol.com/pdfs/TASER_Media_Bulletin.pdf.

TRAC (2008). *TRAC Immigration Judge Reports: Frequently Asked Questions*. Transactional Records Access Clearinghouse, Syracuse University.

— (2018). *Immigration Court Backlog Tool: Pending Cases and Length of Wait in Immigration Courts*. Transactional Records Access Clearinghouse, Syracuse University.

Tversky, Amos and Daniel Kahneman (1973). "Availability: A Heuristic for Judging Frequency and Probability". In: *Cognitive Psychology* 5.2, pp. 207–232. ISSN: 0010-0285.

Tyler, Tom R and Yuen J Huo (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation.

Wasney, Michael (2017). *The Shots Heard Round the City. Are Chicago's new shot detection and predictive policing technologies worth it?* URL: https://southsideweekly.com/shots-heard-round-city-shotspotter-chicago-police/.

Weitzer, Ronald (2002). "Incidents of Police Misconduct and Public Opinion". In: *Journal of Criminal Justice* 30.5, pp. 397–408. DOI: 10.1016/s0047-2352(02)00150-2.

Weitzer, Ronald and Steven A. Tuch (2004). "Race and Perceptions of Police Misconduct". In: *Social Problems* 51.3, pp. 305–325. DOI: 10.1525/sp.2004.51.3.305.

White, Clair, David Weisburd, and Sean Wire (2018). "Examining the impact of the Freddie Gray unrest on perceptions of the police". In: *Criminology & Public Policy* 17.4, pp. 829–858.

Wolfe, Scott E et al. (2016). "Is the effect of procedural justice on police legitimacy invariant? Testing the generality of procedural justice and competing antecedents of legitimacy". In: *Journal of Quantitative Criminology* 32.2, pp. 253–282.

Yousef, Odette (2017). *Gunshots Or Fireworks? How Chicago Cops Know*. URL: http://interactive.wbez.org/everyotherhour/shotspotter/.

Zhang, Zhenbo, Ting Xue, and Xiaoyu Jin (2020). "Effects of Meteorological Conditions and Air Pollution on COVID-19 Transmission: Evidence from 219 Chinese Cities". In: *Science of the Total Environment* 741, p. 140244.

Zhou, Min and Virginia P Sisiopiku (1997). "Relationship Between Volume-to-Capacity Ratios and Accident Rates". In: *Transportation Research Record* 1581.1, pp. 47–52.

Zoorob, Michael (2020). "Do police brutality stories reduce 911 calls? Reassessing an important criminological finding". In: *American Sociological Review* 85.1, pp. 176–183.