COLUMBIA | SIPA
Center for Environmental Economics and Policy

CEEP Working Paper Series
Working Paper Number 24

February 2023

Symptom or Culprit? Social Media, Air Pollution, and Violence

Xinming Du

https://ceep.columbia.edu/sites/default/files/content/papers/n24.pdf

Symptom or Culprit?
Social Media, Air Pollution, and Violence

Xinming Du*

February 15, 2023

Abstract

This paper provides the first causal evidence that hostile activities online lead to physical violence. Given the recently documented relationship between pollution and social media, I exploit exogenous variation in local air quality as the first step to instrument for online aggression. In an event study setting, I find volatile organic compounds (VOCs) increase by 7% when refineries experience unexpected production outages. Together with higher air pollution, I find more aggressive behaviors both online and offline, as well as worse health outcomes near refineries. A one standard deviation increase in surrounding VOCs leads to 0.16 more hate crimes against Black people and 0.23 more hospital visits per thousand people each day. Second, I consider how emotional contagion spreads through social networks. On days with pollution spikes, surrounding areas see 30% more offensive and racist tweets and 12% more crimes; those geographically distant but socially networked regions also see offensive and racist tweets increase by 3% and more crimes by 4.5%. Nationally, overlooking spillovers would underestimate crime effects of pollution by 24%. My findings highlight the consequences of social media hostility and contribute to the public debate on cyberspace regulation.

*Du: Columbia University, email: xd2197@columbia.edu. I am grateful to Douglas Almond, Jeffrey Shrader and Wolfram Schlenker for their continual guidance and encouragement. I thank Maximilian Auffhammer, Vincent Bagilet, Sandra Baquie, Prabhat Barnwal, Elinor Benami, Sandra Black, Tarikua Erda, Caroline Flammer, Ben Gilbert, Florian Grosset, Geoffrey Heal, Sarah Jacobson, Patrick Kinney, Justin McCrary, Jesse McDevitt-Irwin, Tamar Mitts, John Mutter, Suresh Naidu, Waseem Noor, Maya Norman, Nick Obradovich, Cristian Pop-Eleches, Rodrigo Soares, Brigitte Roth Tran, Eugene Tan, Charles Taylor, Eric Verhoogen, Andrew Wilson, Alice Zhang, Yining Zhu and participants of SLab, OSWEET, CESifo Area Conference on Energy and Climate Economics, RES Annual Conference, WEAI Annual Conference, IPWSD and SusDeveR for their helpful comments. I also thank Curley Andrews (U.S. Energy Information Administration), Michael Bailey (Facebook), Johne White (U.S. Environmental Protection Agency) and Gordon Schremp (California Energy Commission) for their help on data, extensive conversations and insights, Victoria Li and Melodie Yuxiu He for technical assistance on social media data, and Solveig Edlund for proofreading. Fundings from the Center for Development Economics and Policy and PhD program in Sustainable Development at Columbia University are gratefully acknowledged.

1 Introduction

Public discourse has become more extreme and polarized in recent years. The UN Strategy and Plan of Action on Hate Speech states that "a disturbing groundswell of xenophobia, racism and intolerance" is happening around the world, and "public discourse is being weaponized with incendiary rhetoric that stigmatizes, dehumanizes and treats minorities as threats" (UN, 2019). Concurrently, social media is an increasingly important venue for public discourse, supplanting traditional forums. Survey results show the majority of the population uses social media as much or more compared to traditional media for news consumption, civic engagement, sharing opinions about social and political issues, and showing support or opposition for public policies (Thomas, 2015). Apart from hostility and polarization, harmful content in cyberspace raises several concerns, including religious hostility and intolerance (Mitts, 2019; Muller and Schwarz, 2020b), propaganda of violent and terrorist groups (Mitts et al., 2022), censorship and misinformation (Vicario et al., 2016; Azzimonti and Fernandes, 2022), political interference (Stella et al., 2018; Almond et al., 2022), and privacy and safety (Jain et al., 2021). Social media content could be a manifestation of underlying trends in polarization and hostility. However, aggressive content and misinformation on social media may also fuel the fire and contribute to additional radicalization and extremism. In this paper, I explore how aggressive content is perpetuated on social media and in turn affects physical violence.

Policymakers and private companies have made halting efforts to police online content. In 2016, the European Commission in conjunction with Facebook, Microsoft, Twitter, and YouTube produced a "code of conduct on countering illegal hate speech online".[1] IT companies have invested millions of dollars in trying to identify and remove suspicious accounts. Despite the wide interest and public debate, empirical evidence on social media and violence is quite limited.

Previous studies have shown the causal effects of online activities on other outcomes, despite violence being the most worrisome and important manifestation of online discontent. For instance, focusing on online emotion spread, Kramer et al. (2014) experimentally deleted happy posts on Facebook and found other users' expressed happiness to be negatively affected. Regarding offline activities, Enikolopov et al. (2016) show social media adoption leads to more public event participation due to easy communication and low coordination cost. A very pressing question is whether and how much social media affects crimes in the real world. This study provides the first causal evidence on how online aggression affects crimes and violence in daily lives.

The challenge of providing a causal answer is to find a plausibly random variation in "grumpy" activities online, or an instrument that increases social media aggression. With the random variation, causal effects capture differences between offline violent responses given different exposure

[1] Unlike the EU, the U.S. government is prevented from punishing or censoring speech by the First Amendment to the Constitution, even many forms that may be considered hate speech. It relies on private companies to police toxic content online on their own platforms, which may affect user experience, raise debates on speech censorship, and hurt company benefits. Despite policy debate, I use the U.S. to conduct my empirical strategy on the impact of online aggressive content with less government censorship concern.

to online aggression. Random assignment ensures that potential crime outcomes would stay the same in the absence of exposure difference. The instrument should satisfy three features. First, it should significantly affect aggressive online behaviors (relevance). Second, the instrument should be exogenous and unexpected. The corresponding increase in aggressive activities should be sharp. Otherwise, it gives online users time to adjust to harmful content, and statistically difficult to remove trends from effects of interest. Third, the instrument should exclusively affect online aggression and not affect crimes of interest or other channels that also affect crimes. Ideally, there should be spatial differences in the online driver and offline outcomes. In this paper, motivated by engineering reports, I devise an instrument that satisfies all these three requirements, so as to study the causal effect of social media aggression.[2] I take advantage of an exogenous variation of local air pollution that induces local users near polluters to write more aggressive content online. Then I study geographically remote but connected online regions' crime responses. The entire causal chain is developed in three steps.

In the first step, I examine air pollution near refinery plants. The sector is chosen due to its unique patterns of production disruption. Refinery plants experience normal and abnormal outages, the former mainly for maintenance purposes. In contrast, abnormal outages are unexpected, harmful, and usually lead to excess air pollutants released to the surrounding atmosphere. Mechanically, excess emissions result from unintended leakage of oil vapors, intended gas venting, and catalytic release. I study pollution response by linking 762 abnormal outages' details to satellite volatile organic compounds between 2014 to 2019. On abnormal shutdown days, air pollution within 20km of plants increases by 7.4%. I confirm the exogeneity by showing no pre-trend in pollution before outages start and the unpredictability of outage schedules using observed covariates. These outage events provide natural experiments for air pollution in refineries' surrounding areas.

The second step focuses on the pollution-induced aggression of local social media users. Existing papers show air pollution affects expressed happiness on social media, measured by sentiment scores (e.g. Zheng et al., 2019; Muntifering, 2021). Though online aggression and its response to pollution have not been studied, it is one type of social media activity and is captured by the text content. Therefore, I expect social media hostility to increase when

[2]Local air pollution is one of the instruments that satisfy these three features. Another potential environmental instrument is local weather fluctuation which is shown to be exogenous after controlling for local seasonality (Dell et al., 2012; Auffhammer et al., 2013). Here I favor local air pollution over weather for four reasons. First, locations of air pollution emissions are known and allow for clear separation of local impacts and social network spillovers. In contrast, weather is spatial correlated and has an uncertain degree of autocorrelation (Auffhammer et al., 2013). I avoid this bias of spatial correlation and focus on air pollution instruments. Second, air pollution events of interest are unexpected and sharp due to anthropogenic emitters and unexpected failure. Weather usually has day-to-day modest variation and predicted change with the help of weather forecasts. Using air pollution shocks, I don't have concerns about anticipation effects and potential behavioral changes in response to weather forecasts. Third, there are many news presses about heat waves especially given the increasing interest in climate change. Oppositely, my pollution emission events of interest are only known to polluters and tend to be underreported, as shown in Section S11. Fourth, there are more policy implications regarding emission externality and pollution regulation from air pollution instruments, while policy implications on weather are less clear and longer-termed. Future works could explore other valid instruments to study local and social-network-based effects. I discuss future research directions in Section 11.

surrounding air quality gets poor.[3] To test this hypothesis, I collect 25 million tweets written by users near polluters and construct two machine learning-based classifiers: offensive and racist. In a reduced form design, I show abnormal outage days see 24% more offensive tweets and 30% more racist tweets. In contrast, there is no change in the total number of tweets, daily income, and outdoor activities. The latter two suggest breathing bad air is the only driver of hostile content. If online hostility is a manifestation of users' feelings and a symptom of poor mental health, my findings are consistent with the documented impact of air pollution on psychological conditions (e.g. Tausczik and Pennebaker, 2010; Gao et al., 2022).

As the third step, I move to non-local areas and evaluate the impact of increased online aggression on offline crimes. The treatment is the exogenously increased online hostility written by local users. The transmission channel is the social media platforms. Among remote social media readers, those closely or loosely connected online with refinery areas have different treatment levels. Following existing papers on online spillovers (e.g. Bailey et al., 2018; Fritz and Kauermann, 2020), I use social connectivity indexes to measure online friendship between each county pair. On pollution spike days, geographically distant but online connected regions also see offensive and racist posts increase by 1-3% and crimes by 4.5%.

My non-local analysis in step three is related to the empirical literature on estimating spillover effects. In a difference-in-difference setting, the stable unit treatment value assumption (SUTVA) is needed to identify the average treatment effect (Rubin, 1978; Manski, 2013). That said, outcomes for a given observation respond only to its own treatment and are invariant to treatment status of others. In a spillover setting, social interactions matter, outcomes vary with the treatment of others, so SUTVA is violated. In my local analysis on pollution impacts, I tackle spillovers and SUTVA concerns through two approaches, one by controlling for inward spillovers, and a second by restricting sample only using online isolated plants. For the first, I find similar responses in local social media and crimes on local event dates, with and without connected regions' events controlled. For the second, I show strongly robust effects using the whole sample and restricted isolated sample.

Another econometrics concern regarding spillover regressions is raised by Borusyak and Hull (2021), who argue that spillover effects are co-determined by two dimensions of variations: exogenous treatments and pre-determined network adjacency. To address the latter non-exogenous component, I simulate counterfactual shocks by reshuffling outage plants and dates to construct placebo treatments. If the remote increase in aggressive behaviors is driven by connected areas' hostility, the degree of exposure (pre-determined connectivity) should not predict an increase in aggressive behaviors at placebo treatment (pollution spike) dates. Otherwise, the estimated effect is driven by the higher aggression in areas at the higher end of the connectivity distri-

[3] According to medical studies, both animal and human activities significantly change and become more aggressive when exposed to poorer air quality. For example, Musi et al. (1994) and Petruzzi et al. (1995) show mice are more likely to fight and defend against other mice when air pollution is higher. Similar evidence is found in monkeys (Chen et al., 2003). Focusing on humans, serotonin levels that are responsible for the brain and nervous system significantly decrease with air pollution and in turn lead to lower impulsive control and a higher likelihood of aggressive behaviors (Coccaro et al., 2011; Crockett et al., 2013). Section 2.3.1 provides detailed literature review for biological mechanisms.

bution. I confirm that this is not the case by showing robust results using actual treatment or recentered treatment.

Empirical findings aside, I model social welfare when environmental externality and social interaction externality are interdependent. Building on the work of Pigou (1920), environmental externality exists when producers equate private cost and benefit of production, ignoring external costs of pollution. I modify the utility function by adding social interaction market. Private benefit occurs when people generate social interaction, like sharing happiness and venting frustrations. The private cost of networking is the time spent on social activities and the lost income. Externality arises when people receive interactions. High quality in the baseline interaction pool means more external benefit, while lower quality generates external cost. I calculate the optimal tax levels on emission and social interaction when pollution worsens social behaviors. I find the second-best emission tax should be higher than the first-best one given the mutual interaction of multiple externalities. Furthermore, my model also contributes to the literature by simulating social contact and its optimal level. Results show both quality and quantity matter in social relationships.

My findings point to the power of social media and its effect on physical violence. I find the increase in distant crimes could not be explained by pollution dispersion, physical movements, or changes in oil product markets. Therefore, I am able to attribute responses in connected regions to online activities and propagation. A crime increase of 4.5% in each county is equivalent to 154 more crimes per day in the whole U.S. I also find significant increases on the following day and no changes afterward or before the pollution spike, indicating no displacement and an absolute increase in the total number of crimes. To make matters worse, non-local crime response is widely distributed across online connected counties. Careful consideration of policies is needed to address such violence increases when concentrated policing efforts may not be helpful.

This paper adds to the economics literature on the causal impact of media platforms on crime by studying online social media, contrary to existing evidence on traditional media. In lab and natural experiments, conventional media platforms like newspapers (Gerber et al., 2009), television (Card and Dahl, 2011), movies (Dahl and DellaVigna, 2009), and the radio (Wang, 2021a) have been shown to significantly affect violent behaviors. This paper complements this area by studying social media and online networks, today's dominant media forms. Compared with traditional ones, social media platforms are being used for longer hours, have more rapid spread, lower costs to participate, and multi-directional feedback. My results on increased crimes are consistent with the persuasive functions of media platforms.

Apart from the causality contribution, my paper extends the literature by providing the first evidence that social media is a culprit for crimes. Previous correlational papers focus on crime detection and prediction using tweets or other social media posts.[4] They mostly use text mining and content analysis to evaluate whether social media data helps to find crime incidents (e.g.

[4]Another aspect of correlational studies focuses on criminal actions committed on social media, like fraud, cyberstalking and harassment, and information leakage. Drury et al. (2022) provides a detailed review of novel crimes created on social media landscapes.

Aghababaei and Makrehchi, 2016; Siriaraya et al., 2019; Sandagiri et al., 2021). Example signals include racial bias, xenophobic messages (Williams et al., 2020), and weapon usage (Fowler et al., 2020). In contrast, whether social media content increases crime as an initial driver rather than a predictor or manifestation has not been studied yet, even in a correlational setting. Furthermore, the second contribution of my paper lies in the spatial difference. Previous correlational studies focus on same-region crime prediction. Researchers draw insights into local crime incidents from nearby tweets. Instead, my paper examines both local and non-local areas' crimes beyond geographic range, and has broad implications on the online network spread.

One strength of this paper is an end-to-end analysis of the whole causal chain. I provide estimates on air quality changes in local areas during exogenous shocks, local effects of online aggression, as well as crime responses in remote regions and their differences with respect to the online friendship matrix. Related to my studies, existing papers are progressing in these three fields by providing partial estimates under different settings, including pollution effects of other natural experiments (e.g. Deschenes et al., 2017; Zhong et al., 2017), impacts of environmental conditions on social media sentiment and crimes (e.g. Burkhardt et al., 2019; Bondy et al., 2020) and emotional spread online (e.g. Kramer et al., 2014). Admittedly, providing answers to some questions is important, but it may take other estimates as given when drawing implications, without checking if earlier settings are generalizable. In contrast, I build a full causal chain, identify effects in each step, and conclude based on my own estimates. In this specific way, my findings have less external validity concerns,[5] and provide policy suggestions to multiple questions.

Another strength of this paper is that I build on the actual social interaction matrix across regions without the need for manipulating connections or online friendships. Nor do I change people's behaviors of engaging with each other or reading more online posts. Instead, I shock the existing endogenous networks and evaluate how people respond. The intervention is not manually assigned to some participants but covers a larger population. Exploring actual social connectedness suggests my findings are close to real-world conditions and more generalizable for policy design. As the treatment of interest - exposure to online aggressive content - is harmful to readers, randomization studies are not possible given the ethical concern.[6] Furthermore, informed consent is a legal requirement for experimental studies and may affect later offline activities of interest (Nijhawan et al., 2013). Manipulating social networks or web traffic is also likely to generate spillovers to the designed control group and flow burdens to online platforms. In addition, my observational setting that aggregates online users at the county level gives me higher power of statistical analysis and enables me to analyze downstream crime effects. It is challenging for experimental studies to track the same users' crime and other behavioral

[5] External validity concerns are more severe when studying air pollution effects as they tend to have nonlinear relationships with downstream outcomes (e.g. Chay and Greenstone, 2003; Bondy et al., 2020).

[6] Kramer et al. (2014) received several ethical concerns after the paper publication, including lacking informed consent from experiment participants, platforms' sharing user network information and feed content to researchers without user permission, short-term emotional damage on participants captured by the paper findings, potential long-term damages that are not measured in the experiment, additional negative effects on participants after reading the paper and interpreting their rights violated (Flick, 2016; Panger, 2016; Hallinan et al., 2020).

responses after the treatment.

The adverse effect of social media aggression raises the need to understand the drivers of online activities. Existing studies show both internal and external factors are predictive of social media activity. Internal factors like personality, outcome expectations, and self-efficacy are associated with online behaviors and the likelihood of posting negative content (Lu and Hsiao, 2009; Yen, 2016). On external factors, most studies focus on significant social events, and they find social media responses quickly to external events like local emergencies, the Black Lives Matter marches, and the COVID lockdowns (Latonero and Shklovski, 2011; Ince et al., 2017; Merchant and Lurie, 2020). This paper contributes to this question from the environmental perspective. Compared with other drivers, environmental factors have sizeable, day-to-day variations that would generate frequent responses online. Moreover, environmental stressors are crucial to an increasing extent under climate change, frequent natural disasters, and severe ecological degradation.

My local first stage analysis contributes to the environmental economics literature by providing an innovative natural experiment for air pollution. Close to my topic, there is substantial literature using weather conditions like wind directions or atmospheric inversions to instrument the air pollution (e.g. Arceo et al., 2016; Deryugina et al., 2019). Also, some papers study a small number of unexpected events or focus on a limited geographic range to identify pollution effects (e.g. He et al., 2016; Lavaine and Neidell, 2017). My identification strategy exploits the unexpected and widespread outage events. The natural experiment is reoccurring over time and is common across refineries worldwide. The variation does not rely on weather conditions, so my study does not have the exclusion restriction concern that weather itself affects air quality and downstream outcomes. I also overcome the low power concern and provide evidence on a common polluting problem.

Also in the area of environmental economics, my study extends our understanding of externalities by quantifying the non-local effects of environmental stressors. Unlike the well-discussed impacts due to exposure[7], how pollution indirectly affects people has not been explored. Another externality exists if exposed people affect other exposed and unexposed groups through online and physical social interactions.[8] In my empirical analysis, I find 12% more actual crimes in local areas and 4.5% more crimes in non-local but online connected areas on each pollution spike day. Taking the whole U.S. together, non-local crime effects are 31.8% as the local effects. My results imply that ignoring the second externality leads to an underestimated external cost of pollution and biased environmental decisions.

On the policy front, this study is the first to uncover and examine the impact of refineries'

[7]Section 2.3 and S2 provide thorough reviews of pollution impacts on surrounding residents. In short, existing studies show that physical health (Schlenker and Walker, 2015; Deryugina et al., 2019), mental health (Zhang et al., 2017), cognitive skills (Nauze and Severnini, 2021), productivity (He et al., 2019), and sleep (Heyes and Zhu, 2019) get worse with poorer ambient air quality.

[8]Theoretically, the second externality could be internalized by regulating social interactions, e.g., imposing a tweet tax or filtering hate speech in text messages. But these practices are costly and hard to capture all channels of social networks. In contrast, addressing the second externality by setting a correct upstream pollution tax may be a cost-effective way.

abnormal operations. While refinery pollution is broadly discussed and has been regulated since 1995, abnormal operations are not informed to the general public or regulated in most states. The severity of air pollution from abnormal operations raises the need for environmental regulations and public awareness. In addition, my findings emphasize the environmental injustice of the refining sector. I find Black Americans encounter not only uneven burdens of pollution exposure but also more hate crimes when air quality gets worse. My results add to a new dimension of environmental injustice.

Another policy implication lies in the debate surrounding cyberspace regulation. Social media gives a new forum to offensive forms of expression. However, any effort to eradicate hate speech runs the risk of expunging legitimate political expression. According to the UN, freedom of expression plays a critical role in promoting equality and in combating intolerance, and the role the media, the internet, and other digital technologies play in keeping society informed is essential (Kaye et al., 2016). This controversy dates back to 2001, when anti-Islamic hate speech sprouted up on the internet. It has been renewed in recent years as the U.S. experiences the Black Lives Matter and Me Too movements that raise consciousness and promote national dialogue about racism and sexual harassment. We also see increased calls for laws punishing speech that is racially harmful or offensive based on gender identity. On all sides of the public debate, we try to seek a balance between fully respecting freedom of expression and keeping harmful speech from escalating into hostility and violence. My study contributes to the public debate on cyberspace regulation by providing neutral and objective causal evidence.

The rest of the paper is organized as follows. Section 2 provides a brief description of related literature, and Section 3 presents conceptual frameworks on multiple externalities and optimal taxation on emissions and social interactions. Section 4 describes the data sources. Section 5 details the empirical strategy. First stage results are presented in Section 6, followed by reduced form results in Section 7 and 8. In Section 9, I outline the empirical strategy, identification threats and solutions, and results of spillover analysis. Section 10 explores potential mechanisms for aggressive behaviors, and Section 11 concludes.

2 Literature review

My study is most relevant to three pieces of literature. First, my study contributes to the literature on social media and real-world violence. Section 2.1 provides a review of psychological theories on media and persuasion, as well as empirical evidence on media impacts. Second, Section 2.2 discusses multiple externalities in environmental economics. When using market-based instruments, other distortions would affect the optimal level of environmental tax. Third, air pollution, together with other environmental hazards, has substantial adverse impacts on human health, crimes, and online behaviors. I review pollution impacts papers in Section 2.3 and S2. In addition, this paper is also related to a handful of studies on the oil refining sector, its environmental impacts, and responses to regulations. Among them, refinery outages are shown to affect downstream oil markets and gasoline prices.

2.1 Social media and real-world violence

2.1.1 Psychology

This paper provides evidence that aggressive content on social media affects physical crimes in real life. Social media is an internet-based form of media platform. As a way of communication, social media, as well as traditional media, provides a channel through which a message or information can be transmitted from the communicator to the recipient. Broadly speaking, my study contributes to a substantial body of psychological literature on media and its persuasion and influence. Theoretically, Felson (1996) provides a systematic review and lists psychological reasons why media affects violent behaviors. First, learning and imitation. People imitate behaviors they see on mass media, and the process of imitation is emphasized by social learning (Bandura, 1983). If imitation and learning include behaviors that involve an intent to harm, exposure to aggression in the media affects the incidence of violence. Second, cognitive priming. Aggressive ideas and content in media platforms could activate other existing aggressive thoughts in viewers through their association in memory pathways (Berkowitz, 1984). Third, emotional arousal. When the media recipient is already prone to act aggressively but instead behaves in some other way, then emotional arousal will facilitate the aggressive behavior (Bandura, 1973). Fourth, sponsor effects. Viewers are likely to believe that the violent presentation is condoned by the media sponsor and the audience (Wood et al., 1991).

Lab experiments examine short-term consequences of media violence exposure. Studies show that people in laboratory experiments who observe media violence tend to behave more aggressively than those in control groups. Meta-analysis of these studies reveals consistent and substantial media effects (Andison, 1977; Martins and Weaver, 2019). To address the external validity of lab experiments, some field experiments are conducted to test the media's large-scaled persuasion effects. They mainly focus on non-violent behaviors, probably due to ethical concerns. For example, Gerber et al. (2011) randomly assigns launch dates and volumes of television advertising and finds significant responses in media buyers' partisanship. Gerber et al. (2009) randomly sends out free press subscriptions and concludes newspaper affects political standpoints and voting behaviors. On crimes and conflicts, a radio soap opera in Rwanda with messages about reducing intergroup prejudice is shown to change listeners' perceptions of social norms. The treated group has more trust, empathy, cooperation, and trauma healing compared with the unexposed group (Paluck, 2009).

2.1.2 Correlational studies

By facilitating the circulation of hate speech, social media is responsible for an apparent increase in xenophobic attitudes and hate crimes (Zhuravskaya et al., 2020). While this paper is not the first one to raise the question whether aggressive content triggers violence, all existing papers provide correlation and prediction-based answers. For example, Aghababaei and Makrehchi (2016) shows historical and real-time social media content could provide signals for crime detec-

tion. They use tweets and automatic data annotation to build a prediction model. Wang et al. (2019) show a strong association between drug abuse-related tweets and crimes in the U.S. In India, Mahajan and Mansotra (2021) use tweets as input and adopt semantic sentiment analysis and neural networks to predict crimes. The precision and recall of the prediction model are over 0.80. Williams et al. (2019) find a consistent positive association between Twitter hate speech targeting race and religion and offline racially and religiously aggravated offenses in London. I contribute to the literature by showing online aggression causes offline violence rather than the other way around and ruling out other potential drivers.

2.1.3 Economics

Closely related to this question, a small number of economics papers have shown the causal effects of social media adoption on crimes. To measure adoption, they either use the number of active users in each region or the number of social media posts in a study period. Enikolopov et al. (2016) develop an instrument for social media penetration, taking advantage of the platform founder's alumni network. Penetration is calculated as the number of active users of a dominant social network in Russia in each Russian city. They find a 10% increase in platform penetration increased the probability of protest participation by 4.6% and the number of protesters by 19% during the 2011-2012 protest movement in Russia, which was triggered by electoral fraud. Using the same instrument and penetration measure, Bursztyn et al. (2019) study the effect of social media usage on ethnic hate crimes and xenophobic attitudes in Russia. Another set of papers takes advantage of specific time periods when social media usage abruptly changes. Muller and Schwarz (2020a) exploit the timing of local internet disruptions in Germany as an exogenous variation in access to social media. They find significantly fewer Facebook posts and fewer anti-refugee attacks. Muller and Schwarz (2020b) evaluate the effect of Twitter usage on anti-minority sentiments in the U.S. Twitter usage is calculated as the number of Twitter users in each county. The authors exploit the South by Southwest festival in 2007 that promoted early Twitter adoption by festival participants. They show that anti-Muslim hate crimes in the U.S. have increased disproportionally in counties with higher Twitter penetration after the start of Trump's presidential campaign.

All the causal evidence focuses on the usage and penetration of social media. They attribute the social media impact to an alternative communication channel, lower coordination costs, and richer information sources. In contrast, this paper evaluates how the content or quality of social media posts affects real-world violence. When the number of posts is not affected, more aggressive and hostile online content leads to more crimes. From the policy perspective, these studies focus on quantity restrictions on social media posts or online platforms. In contrast, my findings on post content and its aggression suggest that targeted quantity control would be effective in reducing violent activities in the real world.

In addition, economists confirm the persuasive function of traditional media and its aggressive content on physical crimes, despite no evidence for social media. Real-world violence is shown

to be affected by television (Card and Dahl, 2011), movies (Dahl and DellaVigna, 2009; Bhuller et al., 2013), broadcast advertisements (Huber and Arceneaux, 2007), radio (Wang, 2021a,b), and internet and broadband diffusion (Gavazza et al., 2018). DellaVigna and Gentzkow (2010) and DellaVigna and La Ferrara (2015) provide detailed reviews on traditional media's impacts on crimes, domestic violence, and racial bias, as well as health, labor, and education outcomes. Compared with traditional media, social media has immediate effects without delay due to press times. Also, it has a low cost to write, read, and edit by the general public, while traditional media is primarily authored by journalists and governors and has a more targeted audience. Furthermore, social media has two-way conversations and feedback, whereas traditional media is mainly unidirectional. The rise of social media, as well as its similarity and difference with traditional media, underscores the importance of empirically testing its persuasion and influence.

2.2 Multiple externalities in environmental decision making

A major objective for environmental decision making is to identify and correct externalities. In the classical analysis of market failure, Pigou (1920) shows that the negative externalities caused by pollution would be internalized by the market if polluters paid a tax equal to the marginal external cost of polluting emissions. This proposition is derived under the assumption of perfect competition and the presence of only one market failure, i.e., in the first-best setting. In reality, the proposition is not optimal with the pre-existing other externalities. One important externality is the income tax. With multiple externalities (income distortion and pollution), the introduction of emission tax raises the price of dirty goods, increases leisure demand due to substitution effects, and provides further disincentive to work. The tax interaction effect could not be eliminated by the revenue recycling effect when environmental tax revenues are recycled through lump-sum or income tax cut (Bovenberg and Goulder, 1996). Therefore, pollution tax is more costly in a second-best setting than it could be in a first-best setting due to multiple externalities, and the double dividend claim is not upheld. When designing tax, optimal pollution tax should be substantially lower than the marginal external cost of pollution (Goulder, 1998).

Another pre-existing distortion lies in market structures. Buchanan (1969) and Barnett (1980) show that a pollution tax should be set lower than the marginal external cost of pollution when the polluting industry is imperfectly competitive. The tax trades off the desire to provide incentives for abatement and the necessity to prevent a greater contraction of output. Since a monopolist in his efforts to maximize profits already chooses to produce less and sell at a higher price than what would result under perfect competition, a pollution tax would lead to market output farther away from social optimum and generate welfare loss. Empirically, Davis and Muehlegger (2010) find the markup in natural gas markets in the U.S. faced by residential customers is equivalent to $55 per ton of carbon tax. The authors conclude that further imposing a carbon tax on natural gas sectors would move consumption in the wrong direction further below the efficient level.

Based on the theory of the second-best and multiple externalities, several authors have ex-

plored, quantified, and refined the conclusions under more specific industry structures and policy settings. Unlike the classical papers, they find the optimal emission tax could be higher than the Pigouvian tax (e.g. West and Williams, 2004; Stavins and Bennear, 2007; Nimubona and Sinclair-Desgagne, 2013; Bento et al., 2014).[9] In this paper, I study how the presence of environmental externality and social contact externality affects the optimal emission tax. Pollution makes people more negative, so negative social externalities are amplified. The marginal external cost of pollution includes the well-studied pollution damage, as well as the distortion in the social connection market. Therefore, the second-best emission tax should be higher than the first-best emission tax.

Compared with classical second-best models, my model adds extra assumptions to the utility function. In contrast, Bovenberg and Goulder (1996) and Buchanan (1969) change assumptions on budget constraints. In terms of the magnitude difference, Bovenberg and Goulder (1996) estimates the second-best tax should be 28%-73% as of the first-best, due to the welfare loss through income taxes. In Section 9, I find the pollution-induced crime increase in non-local areas is 0.013 as that in local areas. As the number of non-local areas is much larger than the number of areas directly exposed to refineries, the estimated total burden is 31.8% of the total burden in local areas in the U.S. Ignoring the impacts on unexposed groups leads to an underestimation of the environmental externality, and the optimal second-best emission tax should be higher than the first-best level by 24.1%.[10,11]

[9] West and Williams (2004) provides empirical evidence that the second-best emission tax should be higher than the first-best one. The authors show positive impacts of gasoline price on labor supply, indicating gasoline is a complement to leisure and gasoline tax could correct environmental externality and income distortion together. The authors conclude that the optimal gasoline tax should be higher than the marginal external cost of gasoline by 35%. Focusing on the market structure, Nimubona and Sinclair-Desgagne (2013) assumes emission abatement is conducted by environment firms and eco-industry rather than polluters themselves. The centralized eco-industry and its market power result in abatement prices higher than the marginal cost of abatement, and the markups dissuade polluters from investing in abatement. To counter this imperfectly competitive eco-industry, the regulator has to play tougher on polluting emissions and impose a higher tax than the marginal external cost of pollution. Another empirical paper on pollution by Bento et al. (2014) evaluates the overall welfare effects of the Clean Air Vehicle Sticker policy in California. The induced congestion cost outweighs its pollution benefits, so the policy generates substantial welfare loss. Regarding information asymmetry, Stavins and Bennear (2007) focuses on labeling programs like EnergyGuide that provides consumers with environmental impacts of each good. When consumers prefer to purchase green goods, externalities exist in both environment and consumption, and correcting information facilitates better decision-making and energy-saving.

[10] Another difference between my model and existing pre-distortion models lies in pollution leakage. One channel of welfare loss in Bovenberg and Goulder (1996) model is the pollution transfer among sectors and regions when emission taxes are put in some of them. I have no assumption on leakage in my model as I focus on social interactions rather than production change and consumption substitution across sectors or regions. In my stylized model, I assume there is only one industry and no trade of goods among regions.

[11] Regarding labor supply, Bovenberg and Goulder (1996) is built on the assumption that labor supply changes with income and consumption expenditure. This is not always true in labor economics, and sometimes labor supply decision is not affected by wage or price, and substitution elasticity and income elasticity could be close to zero (e.g. Kopczuk, 2005; Giertz, 2006; Heim, 2009; Saez et al., 2012). In contrast, my model is not affected by labor participation decisions and is not sensitive to the magnitude of elasticity.

2.3 Impact of air pollution on social media and crime

2.3.1 Biological mechanisms

Existing medical papers provide evidence on why air pollution creases aggressive behaviors. Some animal experiments show that monkeys (Chen et al., 2003; Soulage et al., 2004; Xu et al., 2021) and mice (Musi et al., 1994; Petruzzi et al., 1995; Allen et al., 2013) tend to fight others, increase stress levels and impulsivity in response to air pollution treatment in labs. These aggressive responses to pollution take place with time lags between one hour (Musi et al., 1994) and one day (Xu et al., 2021). The underlying biological mechanisms could be divided into four physical and one psychological reasons. First is the nervous system. Air pollution triggers inflammatory responses in the central nervous system. The neuroinflammation can trigger increased aggression, impulsivity, and depression (Block and Calderón-Garcidueñas, 2009; Beurel and Jope, 2014). Air pollutants are shown to inflame nerve tissues in humans, dogs, and mice (Van Berlo et al., 2010; Levesque et al., 2011). Second is the oxygen intake activity. Air pollution prevents hemoglobin from accepting oxygen. Oxygen deficiency leads to low attention and cognitive impair (Amitai et al., 1998). The third physical mechanism lies in hormone and chemical production. Air pollution decreases serotonin production (Coccaro et al., 2011; Murphy et al., 2013; González-Guevara et al., 2014). Serotonin decreases impulsive controls, lowers harm avoidance, and increases aggressive behaviors (Faustman et al., 1993; Frankle et al., 2005). Closest to my study, gasoline vapors are shown to substantially increase testosterone production in male rats (Uboh et al., 2007). Testosterone is associated with violent crime in humans (Dabbs Jr et al., 1995; Birger et al., 2003). The fourth reason lies in general physical discomfort. Air pollution triggers headaches, pain, and discomfort. Physical discomfort is the central hypothesized mechanism for temperature-aggression relationship (Baron and Bell, 1976; Anderson and Bushman, 2002; Ranson, 2014). Finally, psychological systems respond to air pollution. Air pollution has been shown to decrease cognitive performance, tolerance for frustration, and ratings of other people and the environment (Rotton, 1983; Rotton et al., 1978). Air pollution also triggers depress, anxiety and irritation (Zhang et al., 2018; Rajper et al., 2018; Anderson and Bushman, 2002; Nattero and Enrico, 1996). Given these biological responses, air pollution mechanically makes people more aggressive and more likely to commit crimes.

2.3.2 Empirical papers

Recent economics literature documents that crimes increase with ambient air pollution. Burkhardt et al. (2019) evaluate short-term effects of pollution exposure in the U.S. 2006-2013 by employing high-dimensional fixed effects and addressing confounding weather variations. A 10% increase in $PM_{2.5}$ and ozone is associated with a 0.14% and 0.3% increase in violent crimes, and the effect is mainly driven by increases in assaults. Using data in London 2004-2005, Bondy et al. (2020) show an additional 10 AQI points instrumented by atmospheric inversions and wind direction increase the crime rate by 1.2%, and experiencing an AQI of above 35 leads to 3.7%

more crimes. Pollution has more substantial effects on crimes that are more spontaneous. A potential underlying channel is higher discounting for future punishment on high pollution days. Chen and Li (2020) take advantage of the pollution decrease due to the NO_x Budget Trading Program and show reduced violent and property crimes in participating states by 3.7% and 2.9%. Herrnstadt et al. (2021) uses wind direction as an instrument to assess pollution and crime in Chicago 2001-2012. They find air pollution increases crime on the downwind sides of interstates by 1.9%. A one standard deviation decrease in PM_{10} results in a 2.9% reduction in violent crime. Separating subcategories, aggregated batteries increase while assaults decrease. Jones (2022) uses dust storms as exogenous sources of air pollution variation. Each storm events increases $PM_{2.5}$ by 49.9% and violent crimes by 12.7%.

This paper contributes to the literature on the air pollution-crime effect, albeit in a different context. Moreover, I provide the first empirical evidence that air pollution causes more hate crimes. Compared with other crimes, hate crimes have broader effects as victims include not only the crime's immediate target but also others like them. Hate crime generates greater psychological and emotional trauma in the victim's community (Iganski and Lagou, 2015). My findings have important policy implications, given the surge in bias-motivated incidents in the U.S. and other countries.[12]

Regarding the impact of air pollution on social media activities, the literature is small and growing, probably due to the recentness of social media usage and research. Most studies use expressed sentiment scores as outcome variables. Sentiment analysis is a useful technique to derive users' emotional states and psychological conditions (Tausczik and Pennebaker, 2010; Gao et al., 2022). When pollution worsens psychological conditions, sentiment and online activities serve as manifestations of users' feelings. A small number of papers document the statistical relationship between air pollution and sentiment. Zheng et al. (2019) show one standard deviation increase in the $PM_{2.5}$ (or AQI) is correlated with a 0.043 (or 0.046) standard deviation decrease in the happiness index in China. The effect is more potent on weekends and holidays and for females. Also in China, Tao et al. (2019) shows tourists' expressed sentiment increases hand in hand with air quality improvement over time. Focusing on investor sentiment, Muntifering (2021) documents an association between criteria pollutants and stock returns in New York City. Food products and wholesale portfolio returns increase with negative investor sentiment, while personal services portfolio returns decrease with negative sentiment. The authors use psychological stress and behavioral isolationism to explain the findings. Apart from revealed sentiment, some other papers focus on stated happiness from survey datasets. For example, Zhang et al. (2017) find one standard deviation improvement in the air pollution index creates a 0.03-0.04 standard deviations' increase in mental health. $18\mu g/m^3$ increase in average $PM_{2.5}$ in the past month increases the probability of severe mental illness by 6.67 percentage points (Chen et al., 2018). Similar evidence that air pollution decreases self-reported happiness are also found in Australia

[12]Hate crime incidents in 37 major U.S. cities increased by 39% from 2020 to 2021, with anti-Asian incidents increasing by 224%. Canada reports a 72% jump in hate crime rates from 2019 to 2021. In the U.K., racially motivated crimes increased by 12% in 2021 (Schumann and Moore, 2022).

(Ambrey et al., 2014), U.S. (Levinson, 2012), U.K. (MacKerron and Mourato, 2009; Ozdamar and Giovanis, 2017), Germany (Rehdanz and Maddison, 2008), and other European countries (e.g. Welsch, 2006; Di Tella and MacCulloch, 2008).

This study makes three contributions to the air pollution-social media literature. First, unlike the well-discussed correlation, my design-based econometric analysis provides the first causal evidence of air pollution on social media content and expressed sentiment written by the general public. Second, online behaviors are restricted within sentiment score, which is not a perfect measure of psychological health and is only one dimension of online behaviors. I use both sentiment and two aggression measures as outcome variables. The latter measures are constructed based on machine learning and have good classification performances. They provide more information on toxic and aggressive content than the commonly-used sentiment score. Third, existing papers estimate the social cost of pollution but only study one-dimensional environmental externality. In my paper, there are two externalities from the environment and social interaction. Social interactions generate positive and negative externalities, while pollution causes only negative ones. Correspondingly, there are two equilibriums, the optimal quantity of pollution and the optimal amount of interactions. In Section 3, I show pollution has two-dimensional effects: it makes the socially optimal quantity of pollution lower than private optimal quantity; it increases the proportion of harmful interactions and exaggerates the negative half of social externalities. Compared with existing papers that prove we need an emission tax to internalize the marginal social external of pollution, I show we need to play even tougher on polluters.

2.4 Refinery pollution, outages, and regulation

As a vital process of the oil industry, refining plants process crude oil into useable products such as gasoline, heating oil, jet fuel, petrochemical feedstocks, waxes, lubricating oils, and asphalt. Refineries generally involve three basic steps: separation, conversion, and treatment. Separation is finished in the distillation unit where the liquids and vapors separate into petroleum components based on their boiling points. Heavy and light fractions end up on the bottom and the top of the tower, respectively. In the conversion step, cracking is a commonly used method that uses heat, pressure, catalysts, and hydrogen to crack heavy hydrocarbon molecules into lighter ones. The goal is to rearrange or split molecules to turn low-value fractions into high-octane components. The final step is treatment, where a variety of streams from the processing units are combined and converted into gasoline blends.

The refining process releases pollutants into the air. When converting crude oil into petroleum products, wastes of different kinds are generated when equipment leaks, combustion of fuels, heating of steam and process fluids, and the transfer of products (Oladimeji et al., 2015). Pollutant emissions are severe enough to be observed from space. Shown in Figure S3.1, smoke rose when the Ferndale refinery plant was operating in the left and middle panels, while there was

no smoke signal when the plant was under planned maintenance in the right panel.[13] Major components of these emissions include volatile organic compounds (VOCs), particulate matter (PM), oxides of nitrogen (NO_x), and sulfur dioxide (SO_2).[14] The refining sector contributes to the second largest share of air pollution damages and the largest share of total VOC emissions among all energy sectors (Jaramillo and Muller, 2016). Moreover, refining facilities are primary point sources located close to people. There are 7.8 refining facilities located near each million population in the U.S., while the number of coal-fired power plants is 3.1 per million.[15] Air pollution aside, refineries also raise wastewater, noise, and safety concerns (Wachasunder, 2004; Myers et al., 2010; Loughery et al., 2013).

Existing science literature documents refineries' air pollution by sampling and chemical component analysis. For example, Nakazato et al. (2015) show the installation of an oil refinery increases metal component in the atmosphere in Brazil. Ragothaman and Anderson (2017) provide a review study on refineries' air pollutant sampling. Within a refinery plant, air pollution results from processing, combustion, fugitive, storage, and auxiliary emissions. Another part of science studies constructs emission inventories and discusses the pollution contribution of refineries. For example, Jaramillo and Muller (2016) estimate air pollution damages of energy production in the U.S. The average pollution damage per barrel processed amounts to 2-6% of the refiner acquisition cost of crude oil. In the economics literature, Lavaine and Neidell (2017) show an oil refinery strike in France led to a significant reduction in SO_2, and increased birth weight and gestational age of newborns. Unlike Lavaine and Neidell (2017) on one event, this paper quantifies the pollution contribution across the U.S. in an extended period and also studies the higher emissions from abnormal operations.

Air pollution burden from refineries is unevenly shared by developing countries and poorer people. Like other hazardous facilities, the refining sector also has environmental injustice concerns. According to the EPA, environmental justice refers to "the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies." There are 135 operating refinery plants in the U.S. Shown in Figure 1, they are widely distributed across the country but are mostly located near vulnerable and minority groups (Carpenter and Wagner, 2019a; Williams et al., 2020). Early papers find hazardous waste facilities - including refineries - are disproportionately sited in minority communities (Boer et al., 1997; Graham et al., 1999). Besides, Carpenter and Wagner (2019b) use emission inventory to analyze

[13] Both water vapor and smoke are emitted and released to the air when the refinery is operating. Mechanically, fuel burning generates pollutants, so the plume should not contain only the water vapor. Though we cannot differentiate vapor and pollutants simply from these images, these satellite images confirm the refinery's operation has substantial impacts on the surrounding environment.

[14] These air pollutants emitted by refineries are harmful and generate adverse effects on health, crimes, and other socioeconomic outcomes. I review existing evidence on air pollutants' impacts in Section 2.3 and S2.

[15] These summary statistics are calculated based on the EPA's National Emissions Inventory (NEI) data (https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei). I obtain NEI Facility-level summary for 2011, 2014 and 2017, manually code refinery or coal power plant based on facility description, and calculate the number of polluting facilities in each county. Then I merge these values with county-level population data from the Census (https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html).

environmental justice in the oil refining industry, and find air pollutant emissions are higher in counties with higher unemployment levels. This is inconsistent with the fair treatment rule that "no group of people should bear a disproportionate share of the negative environmental consequences resulting from industrial, governmental and commercial operations or policies."[16] Therefore, regulating air pollution from refineries not only addresses externalities but also improves equality.

Oil refineries experience outages when production is suspended. There are three types of outages: turnaround, planned outage, and unplanned outage. Refinery turnaround is a planned, periodic shutdown of one or more refinery processing units to perform maintenance, inspection, and repair of equipment and to replace process materials and equipment in order to ensure safe and efficient operations (EIA, 2007). It usually takes 3-9 weeks. Similar to turnaround, planned outages are targeted shutdowns scheduled ahead of time. They are less extensive and and usually take 1-2 weeks. Planned outages help to bridge the gap between turnaround intervals (EIA, 2007).[17]

Unplanned outages occur when unexpected events or immediate plant breakdowns take place and require refinery downtime. They often force a refinery to reduce production sub-optimally (Chesnes, 2015). Unplanned outages are generally disruptive since they only allow a short time to plan for the shutdown. Compared with normal operations, oil refineries emit many more pollutants when unplanned shutdowns occur. One reason for the pollution increase is fugitive emission. Both crude oil and petroleum products evaporate and form VOCs. Control and recovery systems are needed to prevent excess VOC emissions, and they cannot operate normally when there are unplanned outages. Another reason is gas flaring and venting. Shutdowns lead to refinery workers flaring, burning, and releasing unprocessed gas to prevent damage to their processing units. Unlike fugitive emissions which are mostly unintended, gas flaring and venting are intended to protect plant safety. Comparing these two practices, venting generates more VOCs, while flaring results in the release of NO_x and carbon dioxide. The other reason is catalytic release. Under normal operations, catalyst moves in the production line. When unexpected upsets occur, this circulation can be disrupted, resulting in a rapid release of catalysts to the environment. In contrast, planned outages require preparation in advance and do not induce pollution spikes.

While air pollution from refineries' normal operations is widely discussed, there is no study quantifying pollution increase from abnormal operations or unplanned outages. On greenhouse gas (GHG) emissions, abnormal operations and equipment malfunction increase CH_4 emissions in the oil supply chain (Alvarez et al., 2018). The authors conduct sample-based estimation and conclude actual emissions should be 60% higher than the EPA inventory estimate. They suggest abnormal operations are largely responsible for the underestimated emissions. Their findings

[16] Detailed definition and explanations on environmental justice could be found on the EPA website: https://www.epa.gov/environmentaljustice/learn-about-environmental-justice

[17] In this paper, I classify outages into planned and unplanned ones. I focus on the unexpected pollution spike that does not happen if the shutdown is well prepared. As refinery turnarounds are scheduled ahead of time, they are included in planned outages in later analysis.

raise the question whether non-GHG air pollutants also increase during abnormal operations.

Existing studies on refinery outages mainly assess their impacts on oil prices. EIA (2007) find oil refining outages decrease product production. Supplies are abundant relative to demand, so refinery outages are likely to have little impact on the product price. Refinery outages increase oil product prices. The magnitude of price increase is more prominent in areas requiring special fuel blends to meet local fuel needs (Kendix and Walls, 2010). Chesnes (2015) show most outages are positively associated with monthly gasoline prices in downstream markets.[18] The effect of unplanned outages is twice as large as the effect of planned outages.

Not in the academic domain, several news articles descriptively document pollution increase when refineries have unplanned outages. For example, CBS in March 2020 reported an equipment failure in Denver: "*Suncor Energy's refinery plant located just north of Denver city experienced an equipment malfunction at around 5:40 p.m. March 17. The incident caused the release of catalyst, a clay-like substance used in the refining process. Visible plumes of yellowish smoke were being emitted from the facility's smokestacks.*"[19]

This paper also contributes to the environmental economics paper on regulating refinery pollution. Studies find pollution regulation and its induced abatement investment increased oil refineries' productivity in LA (Berman and Bui, 2001). Similar evidence by Shadbegian and Gray (2005) show pollution abatement expenditure does not have significant adverse effects on firm productivity in the oil refining sector. In contrast, Sharma (2013) find oil refineries in the U.S. suffered decreases in productivity due to pollution abatement activities 1974-2000. Sweeney (2015) and Burkhardt (2019) find strict environmental policy increases production cost, increases markups, and causes pollution leakage.

3 Stylized model

This section introduces a simple model of J counties, each with a perfectly competitive industry. Extending the work of Pigou (1920), I consider how environmental tax internalizes externalities. For simplicity, I assume that consumers have fixed wage levels and markets clear in a single period. Apart from a good market, consumers decide their input in a market of social interactions. In the first-best setting with only one externality or multiple but non-interactive externalities, I replicate the finding that the optimal pollution tax equals the marginal external cost of pollution on exposed people. Then I introduce interactive externalities by allowing for pollution's impacts on the quality of social contact. The presence of interactive externalities and knock-on chains leads to pollution's adverse effects on other exposed people or unexposed non-local people. I solve for the optimal pollution tax which is higher than the local estimate in order for the social optimum to be reached.

[18] Given Chesnes (2015)'s findings, oil price could be a potential confounder of my pollution-social media, pollution-crime, and social media-crime story. Section 9.5 shows negligible effects on daily oil prices when abnormal shutdowns and rules out this confounding channel.

[19] Source: https://denver.cbslocal.com/2020/03/18/suncor-equipment-malfunction-yellow-smoke-commerce-city/

Furthermore, I present an alternative model in Section S1 by extending the empathetic preference model by Heal (2021). Social contact does not directly affect utility, but shows up as weights that consumers place on others' utilities in their welfare functions. Environmental externality affects social externality by changing how people care about others. I reach the same conclusion that tougher environmental regulations should be played with the interaction of multiple externalities.

3.1 Model setup

Consider a representative consumer in county i purchases final good of quantity x_i and gains utility of $U_i(x_i)$. Production of local producers generates pollution of P_i, and the marginal damage of pollution exposure is d. Apart from good consumption, consumers also gain utility from social interactions. They benefit from talking to people and writing posts on social media as they share happiness or release anger. Assume the number of social interactions conducted by the representative consumer is s_i, the marginal benefit of social interaction is α, and the total number of interactions by all residents in county i is S_i.

Moreover, listening to others and reading online posts affect utility. The number of social contact received by consumer in county i include local contact and non-local contact in county j. I use C_{ij} to measure the connectedness between county i and j, with $C_{ii} = 1$, $0 \leq C_{ij} \leq 1$. For each consumer in county i, a total of $\sum_{j=1}^{J} S_j \cdot C_{ij}$ interactions come into view. Each consumer's utility maximization is given by:

$$\max_{x_i, s_i} U_i = U_i(x_i) - d \cdot P_i + s_i \cdot \alpha + \sum_{j=1}^{J} [S_j \cdot C_{ij} \cdot (1 - n_j) \cdot \beta_1 - S_j \cdot C_{ij} \cdot n_j \cdot \beta_2]$$

$$s.t.\ I_i + w_i \cdot [T - t_i(s_i)] \geq p_i \cdot x_i$$

where n_j denotes the proportion of negative interactions in county j. β_1 and β_2 are the marginal effect of receiving positive and negative ones respectively, $\beta_1, \beta_2 > 0$. I_i denotes non-labor income, and w_i and T denotes wage rate and labor time endowment respectively. $t_i(s_i)$ is the time spent on social networks. It is bounded below t, and is an increasing and convex function due to the diminishing marginal productivity, i.e. $0 \leq t_i(s_i) \leq T$, $t_i'(s_i) > 0$, $t_i''(s_i) > 0$.

Each consumer maximizes utility taking P_i and S_i as given. Consequently, private optimal x_i and s_i satisfy:[20]

$$U_i'(x_i) - p_i = 0$$

$$\alpha - w_i t_i'(s_i) = 0$$

Private optimum suggests consumers consume goods where the marginal benefit equals the marginal cost, and participate in social activities where the marginal benefit of networking equals

[20] Consumers maximize utility taking pollution and the total number of social interactions as fixed. Each consumer's consumption and networking decision has very little impact on the overall level of pollution and social interactions, so P and S are treated as fixed and therefore do not affect private optimization.

the marginal cost (i.e. forgone labor income).

The production cost of delivering goods is given by $c_i(x_i)$. The industry is perfectly competitive, so firms are also price takers. Each firm in county i offers a quantity of x_i that maximizes the profit function:

$$\max_{x_i} \ \pi_i = p_i \cdot x_i - c_i(x_i)$$
$$p_i - c_i'(x_i) = 0$$

Taking consumers and producers in all counties together, social welfare and its optimization is specified as follows:

$$\max_{\substack{x_1,x_2,\ldots,x_J;\\ s_1,s_2,\ldots,s_J}} W = \sum_{i=1}^{J}\{U_i(x_i) - d\cdot e\cdot x_i + \sum_{j=1}^{J}[S_j\cdot C_{ij}\cdot(1-n_j)\cdot\beta_1 - S_j\cdot C_{ij}\cdot n_j\cdot\beta_2]$$
$$+s_i\cdot\alpha + I_i + w_i\cdot[T - t_i(s_i)] - c_i(x_i)\}$$
$$\Rightarrow \quad \frac{\partial W}{\partial x_i} = U_i'(x_i) - c_i'(x_i) - d\cdot e - (\sum_{j=1}^{J} C_{ij})\cdot\frac{\partial n_i}{\partial P_i}\cdot S_i(\beta_1+\beta_2)\cdot e = 0$$
$$\frac{\partial W}{\partial s_i} = \alpha - w_i t_i'(s_i) + (\sum_{j=1}^{J} C_{ij})\cdot[(1-n_i)\beta_1 - n_i\beta_2] = 0$$

3.2 Non-interactive externalities

First, I assume there is no interaction between social network externality and pollution externality, i.e. $\frac{\partial n_i}{\partial P_i} = 0$. The socially optimal quantity of good production is:

$$\frac{\partial W}{\partial x_i} = U_i'(x_i) - c_i'(x_i) - d\cdot e - (\sum_{j=1}^{J} C_{ij})\cdot\frac{\partial n_i}{\partial P_i}\cdot S_i(\beta_1+\beta_2)\cdot e$$
$$= U_i'(x_i) - c_i'(x_i) - d\cdot e = 0$$

This condition cannot be achieved, as private optimum generates marginal external benefit $U_i'(x_i)$ equals marginal private cost $c_i'(x_i)$. Private markets do not produce Pareto efficient outcomes because firms do not take into account the external cost of pollution $d\cdot e$. To reach social optimum, we create a market for pollution by imposing an emission tax T_i. New private optimum faced by firms in county i is given by:

$$\max_{x_i} \ p_i\cdot x_i - c_i(x_i) - T_i\cdot e\cdot x_i$$
$$p_i - c_i'(x_i) - T_i\cdot e = 0$$

To reach $U_i'(x_i) - c_i'(x_i) - d\cdot e = 0$, we need $T_i = d$. In other words, without interaction with other externalities, optimal Pigouvian tax equals the marginal external cost of pollution. Even if counties are closely connected, and participating in social interactions generates positive or negative externalities, as long as there is no interaction between multiple externalities, the

first-best Pigouvian tax is sufficient to solve the pollution externality.

In terms of the second externality on social networks, $\alpha - w_i t_i'(s_i) = 0$ is the same as private optimum. We need government intervention to correct the last term whose sign is ambiguous:

$$\begin{aligned}\frac{\partial W}{\partial s_i} &= \alpha - w_i t_i'(s_i) + (\sum_{j=1}^{J} C_{ij}) \cdot [(1-n_i)\beta_1 - n_i\beta_2] \\ &= (\sum_{j=1}^{J} C_{ij}) \cdot [(1-n_i)\beta_1 - n_i\beta_2] = 0\end{aligned}$$

A networking tax is needed if $(1-n_i)\beta_1 - n_i\beta_2 < 0$, $n_i > \frac{\beta_1}{\beta_1+\beta_2}$. That said, we need to lower the number of interactions in unfriendly counties with too many negative activities, while in friendly counties, people are not providing enough social interactions. The optimal social interaction tax satisfies:[21]

$$H = max\{0,\ -(\sum_{j=1}^{J} C_{ij})[(1-n_i)\beta_1 - n_i\beta_2]\}$$

where H increases with connectedness, the marginal benefit of good connections, and the marginal cost of involving negative interactions. The policy implication of this result is a potential quantity (rather than quality) regulation of social interactions. Taking online speech as an example, sensitive and hateful content is sometimes monitored, regulated, or removed by mainstream social media platforms like Facebook, Twitter and Youtube. This practice is costly to implement and has some delays (MacAvaney et al., 2019). It also raises debate on censorship and freedom of expression (Howard, 2019). My model shows that an alternative and realistic policy is regulating the quantity of social interactions,[22] and taking advantage of the unobservable market force to reach the social optimum.

3.3 Interactive externalities

Next, I explore how externality interaction affects the optimal pollution tax. I assume $\frac{\partial n_i}{\partial P_i} > 0$, i.e. the quality of social interactions gets worse together with local pollution. The new optimal pollution tax should be:

[21] The first-order condition on s_i suggests governments should subsidize networking activities in friendly counties. However, given the disincentive to work, governments consider social contact a leisure activity and tend not to subsidize it. Therefore, I assume the tax should be either zero or positive.

[22] There are some real-world examples of social interaction quantity control. For example, suspicious accounts posting harmful content are blocked and removed from social media platforms. All their posts are deleted whether negative or not. Besides, Twitter flags potential bots and spammy activities based on quantity. Users are considered spam if sending high-volume messages, posting nearly identical tweets, using excessive hashtags, or engaging with tweets to drive traffic and attention. Moreover, Twitter avoids abusive behaviors by penalizing too-active users' popularity. After a user reaches a certain limit, his later tweets will not be seen by everyone. My model implications show these quantity control measures could improve social welfare, especially when negative accounts are disproportionately targeted. Twitter's spam policy could be found here: https://help.twitter.com/en/rules-and-policies/platform-manipulation Though tweet limits are not specified or disclosed to the public, relevant policies are listed here: https://help.twitter.com/en/safety-and-security/tweet-visibility

$$T = d + (\sum_{j=1}^{J} C_{ij}) \cdot \frac{\partial n_i}{\partial P_i} \cdot S_i(\beta_1 + \beta_2)$$

We need a higher Pigouvian tax than that in the first-best case. While the optimal tax increases with connectedness, it is still larger than d even if counties are not connected. The optimal pollution tax increases with connectedness, and is still larger than d if counties are not connected: $\forall i \neq j$, $C_{ij} = 0$, $T = d + \frac{\partial n_i}{\partial P_i} \cdot S_i(\beta_1 + \beta_2)$.

Focusing on the second externality, a less intuitive result is that the optimal social network tax is not affected by the responsiveness of negative interactions to pollution ($\frac{\partial n_i}{\partial P_i} > 0$). n_i is considered exogenous when deciding H.[23,24] In unfriendly counties, imposing a network tax helps correct pollution externality by reducing S_i. However, if a correct social tax is identified, combining a social tax and a first-best Pigouvian tax could not eliminate the additional pollution externality. Raising the Pigouvian tax is the only solution to the externality interaction.[25] In the following sections, I empirically estimate $\frac{\partial n_i}{\partial P_i}$ at the local scale and its spillover to distant but online connected areas.

Additionally, this model is generalizable to other settings when multiple externalities are interdependent. For example, if I still focus on environmental stressors as the first externality, and the second externality lies in public health concerns, like the COVID spread. Social interactions lead to negative effects due to higher risks of COVID infections, and pollution increases COVID spread risks (Austin et al., 2020; Persico and Johnson, 2021). In this setting, we need a higher emission tax than the local marginal damage to correct the public health externality. The second-best emission tax is even more important when lockdown or social distancing is not easily implemented or enforced.

4 Data

4.1 Refinery plant information

Refinery plant information is obtained from the U.S. Energy Information Administration (EIA). EIA form 820 'Annual Refinery Report' documents each plant's location, capacity, and operator.[26] I use survey 2020 to construct my sample. There are 135 refineries in 12 districts in the

[23] In a dynamic setting where P_i increases over time and n_i increases with P_i, social connection regulators figure out n_i first and choose H accordingly.

[24] I assume n_i changes linearly with P_i, i.e. $\frac{\partial n_i}{\partial P_i}$ is a constant. This may not be true in reality. The marginal cost of air pollution is documented to increase with ambient pollution level (e.g. Muller and Mendelsohn, 2009; Hsiang et al., 2019). The nonlinear effect on n_i does not affect the optimal H.

[25] Nothing about pollution, when a big event occurs and serves as a global negative shock on emotion and social networks, n_i increases in all J counties, so H is higher globally due to the event. Assuming linear pollution effects on social negativeness, the optimal T is not affected by the event and is still higher than the Pigouvian tax. Assuming nonlinear effects, i.e. $\frac{\partial n_i}{\partial P_i}$ increases with n_i, the optimal T is higher globally due to the event. Despite being off-topic in this paper, global negative events are increasingly common these days, like the COVID pandemic since 2020 and the Russia-Ukraine crisis in 2022. My findings show negative externalities in social networks lead to the Pigouvian tax even far from optimal.

[26] EIA form 820 could be downloaded here: https://www.eia.gov/petroleum/refinerycapacity/

U.S. Figure 1 Panel A shows the locations of operating plants. States with the largest number of plants are Texas (30), Louisiana (18), and California (16).

4.2 Abnormal and normal outages

I compare air quality when no outage occurs with air quality when production is suspended to identify the impact of refineries' normal operations. I use temporary outage data from Refinitiv. The dataset provides planned outage and unplanned outage information for all operating plants in the U.S. 2014-2019. It records plant coordinates, outage reason, start and end date, total and offline capacity. I use plant name and location to merge plant information in EIA from 820 and Refinitiv data. The sample ends up with 101 refinery plants existing in both datasets.

Another question of interest is the pollution response to abnormal refinery operations. Refinitiv data allows me to answer this question by checking pollution when unplanned outages and pollution when normal operations.

Figure S3.3 Panel A shows the distribution of planned and unplanned outage start times over the week. Panel B shows the distribution of start month over the year. Planned outages in blue bars are more likely to start on weekdays and in non-summer months. As discussed in Section 2.4, planned shutdowns are targeted and scheduled in advance to perform maintenance and ensure efficient operations. Planned shutdowns tend to happen when energy demand in downstream markets is low. When it comes to unplanned outages, they are still more likely to happen on weekdays but are more evenly distributed than planned events. While unexpected equipment failures could happen anytime and are quasi-randomly assigned across the week and month, it still requires workers to decide and perform the shutdown. Consequently, unplanned outages show a weekly cycle but take place more evenly across time than planned events.

Table S4.2 and Figure S3.3 Panel C show the duration of planned and unplanned outages. Planned outages have a longer duration, with an average of 31 days. Unplanned outages usually last within one week, and the average duration is 12.5 days.

I use refinery production data to verify the outage schedule and assess the direct effect of shutdowns on oil product supply. Refinery net production is provided by the EIA at the district-month level.[27] In Figure S3.4 and Table S4.3, the number of outage events and production have a strongly negative correlation, which suggests outages disrupt refinery operation.

4.3 Air pollution

Air pollutant of top interest is VOC. Common VOCs emitted by human activities include formaldehyde (HCHO), ethylene glycol, methylene chloride, tetrachloroethylene, toluene, xylene, and 1,3-butadiene. Among them, HCHO is one of the most abundant hydrocarbons in

[27] Microdata at the plant-month level is not available. According to the EIA, "individual-level data is protected under CIPSEA (Confidential Information Protection and Statistical Efficiency Act) and cannot be released". Therefore, I use aggregated data at the district-month level to test the impact of outages on production, and aggregated data is downloaded from the EIA website: https://www.eia.gov/dnav/pet/pet_pnp_refp2_dc_nus_mbbl_m.htm

the troposphere and is a secondary pollutant produced by the oxidation of VOCs (Houweling et al., 1998). Existing studies show benzene, ethene, propene and alkenes could rapidly react to produce HCHO (Parrish et al., 2012; Fried et al., 2020).

I use NASA's Ozone Monitoring Instrument (OMI) HCHO product to measure VOC near refinery plants. The product is commonly used to measure ambient VOC levels (e.g. Millet et al., 2008; Zhu et al., 2014; Jin et al., 2020). It has a spatial resolution of 0.1° and a temporal resolution of one day. I use average HCHO density within 20km buffer as HCHO near each refinery plant.

Ambient monitors by the EPA measure benzene and other criteria pollutants. However, the benzene monitors are not close to refinery plants and are not sufficient to capture plants' surrounding VOCs, shown in Figure S3.2. Table S4.1 reports the correlation test between benzene and satellite HCHO near monitors. The significant and positive estimates confirm the validity of the OMI HCHO product.

Table 1 provides summary statistics of average HCHO on outages and working days. Surrounding HCHO is 8.1 units (10^{15} molecules/cm^2) when plants are under normal operations. HCHO decreases to 7.6 units when a planned shutdown happens, and there is no difference on the shutdown day and later days. During unplanned outages, there is a pollution spike on unplanned day 1, with HCHO increasing to 9.1 units. Then it decreases to 8.1 units between the second day and the last outage day.

4.4 Weather

Weather data includes temperature, precipitation and wind speed. The former two variables are derived from PRISM and processed by Schlenker and Roberts (2009).[28] I use the simple average of the daily maximum and minimum temperature to calculate the daily temperature. Wind data is obtained from NCEP NARR.[29] Average wind speed is calculated as the square root of the sum of U-wind and V-wind. I use the average values near 20km buffers of each plant to show surrounding weather conditions.

[28]PRISM data could be found here: http://prism.oregonstate.edu. Daily data for Contiguous United States are obtained here: http://www.columbia.edu/ ws2162/links.html

[29]I use the V-wind at 10 meter Daily Mean product and the U-wind at 10 meter Daily Mean product: https://psl.noaa.gov/data/gridded/data.narr.monolevel.html

4.5 Social media and crime outcomes

4.5.1 Tweet

I collect 25,398,177 tweets within 20km of the top 10 refinery plants 2014-2019 using Twitter API v2.[30,31] Five outcomes of interest include the sentiment of all surrounding tweets, the number and the proportion of air pollution-related, health-related, offensive, and racist tweets. The first variable shows people's general happiness and mood. The second captures people's complaints about pollution and serves as a validation of the pollution change. The third variable measures people's physical health and defensive awareness. The last two outcomes indicate the degree to which surrounding residents are inflamed, aggressive, or out of control. Appendix Section S13 provides the lists of air pollution keywords and health keywords, as well as example tweets in the second to the last outcomes.[32]

Sentiment analysis converts tweet text into a sentiment score that is representative of its emotion. Sentiment in this study is measured based on the Valence Aware Dictionary for sEntiment Reasoning (Vader). This metric is shown to perform well in the social media domain (Hutto and Gilbert, 2015) and is commonly used in existing computer science (e.g. Shelar and Huang, 2018; Newman and Joyner, 2018; Mustaqim et al., 2020) and economics papers (e.g. Baylis, 2020; Almond and Du, 2020). Vader is a lexicon- and rule-based sentiment analysis tool that can handle words, abbreviations, slang, emoticons, and emojis commonly found in social media (Hutto and Gilbert, 2015). Each body of text produces a vector of sentiment scores with negative, neutral, positive, and compound polarities. The negative, neutral, and positive polarities are normalized between 0 and 1. The compound score can be considered an aggregate measure of all the other sentiments normalized between -1 and 1. [33]

[30]Twitter Academic API has a monthly tweet cap usage of 100,000,000 tweets. Collecting all tweets near 20km buffers of 10 refinery plants requires four months' quotas. It is both time consuming and computationally expensive to collect and analyze tweets near all these 101 refinery plants in the U.S. Therefore, I use the top 10 plants with the largest number of unplanned outages for a pilot analysis on tweet-related outcomes. I keep nationwide analysis on other outcomes to show the commonness of outage problems and to maintain sufficient statistical power with a good number of outage events and connected areas.

[31]Collecting geolocated tweets using Twitter API ends up with a sample of Twitter users that turn on geolocation tags. One question is whether this subsample is representative of the general public. On the one hand, the offensive activities of this subgroup should be positively correlated with the general public's activities. On the other hand, I do a user-level analysis by collecting bad authors' followers' tweets and testing their responses to followees' offensive tweets in Section 9.2. This group of followers does not necessarily enable geolocation tags in their tweets, and I still find they post more offensive tweets when their followees are offensive. This supplementary result confirms that the general public have similar responses when observing offensive tweets as the subsample in the main analysis.

[32]One concern is the validity of using social media as proxies for mental health conditions. Existing papers show air pollution negatively affects mental health (e.g. Zhang et al., 2017; Chen et al., 2018). My findings on a higher proportion of offensive tweets and more impulse spending in Section 7.1 and 10.1 are in line with earlier evidence, and could be interpreted as another measure of mental health. Also, it is hard to find perfect measures of mental health. Due to the recentness of mental health care, health outcomes described in Section 8.1 to 8.3 capture pooled effects and focus on physical health. Existing papers on mental health use self-reported happiness or surveyed psychological conditions as outcome variables, which have substantial uncertainty due to sample selection and self reporting bias. My measure is based on expressed posts and revealed evidence on social media, which is at least as objective as other commonly used measures.

[33]I assess whether pollution- or health-related tweets are in general more negative compared with other weets without keywords. Table S7.2 shows a negative and significant estimate on $Pollution\_dummy$ and $Health\_dummy$. People are unhappier by 0.03 standard deviation when talking about air pollution. The magnitude of sentiment score decrease is larger when health-related keywords are mentioned, 0.4 standard devi-

In terms of the last two outcomes, I use supervised learning to classify tweets into offensive and non-offensive ones, racist and non-racist ones. As a first step, I randomly sample 20,000 tweets from all collected tweets, and ask Turk workers to make offensive and racist annotations. Among these 20,000 tweets, 7.1% are labeled as offensive tweets, and 0.86% are labeled racist. Then I use Global Vectors for Word Representation (GloVe) to obtain vector representations for each tweet. GloVe is a pre-trained vector embedding that maps words into a vector space where similar words cluster together and different words repel (Pennington et al., 2014). It is well used in existing studies for text classification (e.g. Wang et al., 2017; Stein et al., 2019). I use GloVe's Twitter dictionary which includes 200-dimensional vectors trained in over 2 billion tweets and 27 billion words. Other dictionaries like Wikipedia and Common Crawl are also used but end up with higher training loss and validation loss. After embedding, I build an eXtreme Gradient Boosting (XGBoost) model to classify 14,000 input tweets (70% of the total labeled tweets). XGBoost has the advantage of self-correcting mistakes (by giving more weights to earlier mistakes and updating weights after later iterations) and accurate loss functions (with a second-order Taylor expansion on the objective function) (Guo et al., 2019). I use the remaining 60,000 to test the model. The training loss and the validation loss for offensive classification are 0.12 and 0.18. The same for racist classification are 0.036 and 0.039. Using the testing set with 60,000 tweets, offensive classification shows an accuracy score of 0.866, F1 score of 0.424, precision score of 0.305, recall score of 0.699, and area under the receiver operating characteristic curve (ROC AUC) of 0.788. Racist classification has accuracy score equal 0.970, F1 score 0.179, precision 0.117, recall 0.380, and ROC AUC 0.677. Finally, I use the trained model to predict 25,378,177 unlabeled tweets. Prediction results show that 7.6% of surrounding tweets contain offensive content and 0.80% contain racist content.

Figure S3.13 displays nouns that appear in offensive tweets and non-offensive tweets. The font size indicates the frequency of each word. The figure suggests that most of the offensive tweets are related to personnel or swear words. Removing swear words, word clouds in Figure S3.14 show offensive and non-offensive tweets have similar topics related to daily lives (e.g. "people", "life", "today", "time", "night", "work", "game", "friend"). For racist and non-racist classification, Figure S3.15 displays balanced topic profiles for two tweet groups. This suggests it is not the change in topics but the usage of hostile words and negative mood that drives the classification and sentiment results. Table S7.3 reports correlation tests of sentiment score, offensive dummy, and racist dummy. In Panel A, tweets with offensive or racist content have lower sentiment scores than normal tweets by 0.76, equivalent to 2.04 of the standard deviation. Estimated coefficients stay robust with alternative fixed effects added, which implies offensive content and sentiment score are not driven by covariates varying across places or times but mechanically correlated. Sentiment aside, offensive tweets are more likely to be racist than non-offensive tweets by 0.06, 4 times the standard deviation of the racist dummy.

Table S7.1 shows these 10 refinery plants' basic information and summary statistics of sur-

ation. T-test shows significant differences in scores whether air pollution is mentioned or not, with t-statistics of -11.16 and p-value of 0.000. Test statistics for health dummy are -46.07 and 0.000.

rounding tweets. They are located in Texas (3 plants), Louisiana (2), California (1), Kentucky (1), Oklahoma (1), Pennsylvania (1), and Washington (1). Eight of them have capacities above 100,000 barrels per day (bpd) and have been operating for over 50 years. The average number of tweets on each day near each plant ranges from 11 to 3322. The average sentiment of nearby tweets is mostly positive, except Monroe Energy Training and HollyFrontier Tulsa where sentiment is close to zero. Besides, air pollution-related tweets account for 0.07-2% of the total, and health-related tweets account for 0.02-0.13%. There are 1.3-418 offensive tweets and 0.1-65 racist tweets near each plant on each day.

4.5.2 Hate crimes and other crimes

Hate crime data is obtained from FBI Hate Crime Statistics.[34] It includes self-reported crime events from Summary Reporting System (SRS) and National Incident-Based Reporting System (NIBRS).[35] Each event is recorded with the incident date, location, the number of victims, offenders, victim's perceived race, and offender's race. Locations are recorded as city or county.[36] I collapse incident-level data into county-day level variables. Outcomes in this paper include the number of hate crime events against different race groups 2014-2019.

NIBRS also provides incident details for other 55 types of crimes about their offenders, victims, offenses, property, and arrest. Compared with hate crimes, this data has lower coverage without SRS' records. Locations in this dataset are recorded as agencies. I use agencies' counties to proxy crime events' locations and to further merge with refinery plants' locations. Outcomes of interest include the number of events of each crime type at the county-day level 2014-2019.[37,38]

Apart from the national data source, there are 14 U.S. cities providing 911 call records for police service (McCrary and Sanga, 2021). They are large cities not close to refinery plants, and I could not use this data for local pollution-crime analysis. Also, the advantage of using NIBRS data is its comparability across counties and years and consistent crime type classifications. To address NIBRS' potential data quality concern, I use the alternative 911 data in non-local crime

[34]Data is available here: https://crime-data-explorer.fr.cloud.gov/pages/downloads

[35]Not all hate crimes are reported by victims (Pezzella et al., 2019). At the same time, the FBI relies on local law enforcement agencies to identify and report crimes (details about the FBI's interaction with other agencies are available at: https://www.fbi.gov/about/faqs/what-is-the-fbi-doing-to-improve-its-interaction-with-other-federal-law-enforcement-agencies), so the coverage of hate crime reporting systems is not complete and may change over time. In my analysis, I am interested in the daily variation of hate crime events within counties, so potential underreporting will be controlled by county and time fixed effects added in the estimation.

[36]I conduct county-level analysis on hate crime and other downstream outcomes without precise coordinates.

[37]Figure S3.9 and S3.10 show county coverage of hate crime and non-hate crime data. The number of counties reporting hate crimes is 1587 in 2014-2019, and is 750, 772, 754, 845, 866, and 951 in each year. That for non-hate crimes is 1,927 in the six-year period, and is 1648, 1673, 1690, 1713, 1792, and 1897 in each year. Figure S3.11 and S3.12 display similar spatial distribution and event count 2014-2019. While county coverages are mostly stable over time, I add county by year fixed effects to address the potential concern of inconsistency reporting. Table S9.22 Panel A reports local hate crime results. Panel B and C display non-hate crime results in local and non-local areas. Results remain stable after controlling for county-year fixed effects. This indicates that crime data coverages are not likely to affect my results in both local and non-local analysis.

[38]I use all counties that report at least one hate crime or one non-hate crime as an alternative sample and consider zero events for non-reporting county-days. This results in larger sample sizes and the same sample sizes for hate and non-hate crimes. Results displayed in Table S7.18 show similar pattern as Table 4. Air pollution induces non-hate crimes and hate crimes against Black people. Estimates on *UnplannedShut* have smaller magnitudes and lower precisions due to more zeros added in dependent variables.

analysis in Section 9.3.

4.6 Health outcomes

4.6.1 Consumption expenditure

I use medical expenditure as a proxy for health conditions. Consumption expenditure data comes from the Nielsen consumer panel. It comprises a representative panel of households that continually provide information about their purchases in a longitudinal study. Nielsen panelists use in-home scanners to record all of their purchases. Consumers provide information about their households and what products they buy, as well as when and where they make purchases.[39] I collapse household-purchase level data into the county-category-day level. Products of primary interest are medical-related (especially respiratory-related) items including cough remedies, sinus remedies, and breathing aids.

4.6.2 Foot traffic

Another health outcome of interest is hospital visits, captured by phone-based movement patterns. I use Safegraph foot traffic data to measure where and how often people move. Safegraph aggregates anonymized location data from numerous mobile applications to capture the movement of people to each point of interest (POI).[40] There are 32.1 million visits to 3.9 million POIs on each day 2018-2019.[41] Outcomes of interest include the number of visits to health and outdoor leisure-related POIs, and overall county busyness.

4.6.3 Mortality

Finally, I test the mortality response to refinery pollution shocks. Mortality data is obtained from Medicare records. I use the Master Beneficiary Summary File 5% sample 2014-2019 that includes 3,029K beneficiaries every year. The dataset contains each beneficiary's county, sex, age, and race. The date of death is available for decedents. From this dataset, I build two outcome variables - mortality and mortality rate[42] - at the county-day level.

[39] Nielsen consumer panel is used in earlier studies to assess consumer behavior. Most studies use food and beverage-related products. For example, Ng et al. (2018) show WIC food package revisions are associated with improved nutritional profiles of food purchases. Households purchase more encouraged food and less discouraged ones consistent with dietary guidance. Janssen and Parslow (2021) find the first pregnancy reduces household alcohol purchases by 36% during the pregnancy and by 34% after pregnancy, but no effect during the second pregnancy. Some studies also use Nielsen data to assess cigarette, sleep aid and drug consumption. Medicaid expansions reduced cigarette consumption and increased smoking cessation product use (Cotti et al., 2019; Wang et al., 2021). People substitute recreational cannabis for conventional sleep medications. The market share of sleep aids declines with higher county-level cannabis sales (Doremus et al., 2019). Austin et al. (2018) focus on weight-loss related products. They find a 20% higher price due to a sale tax is associated with a 5.2% lower purchase.

[40] Safegraph data is widely used to analyze the COVID pandemic, including the spread risk of the COVID (Yang et al., 2020), the social activity response to COVID stringency and reopening (Farboodi et al., 2020; Chang et al., 2021), the spillover effect of anti-pandemic policies (Holtz et al., 2020a; Cook et al., 2020), etc.

[41] Safegraph data is not available 2014-2017, so I only explore movement patterns 2018-2019. Details about the dataset could be found here: https://www.safegraph.com/

[42] Mortality rate is defined as the number of deaths divided by the number of beneficiaries existing in the annual summary file in that year. There are 121K deaths every year, 4% of all beneficiaries.

5 Empirical strategy

I estimate the impact of oil refinery's abnormal shutdowns and normal operation on surrounding air quality by exploring outage schedules. Temporary events are recorded with exact coordinates. I use air pollution near 20km buffers of each plant to estimate the pollution impact of temporary outages. This allows me to capture nearby pollution precisely with the same buffer size for all plants. I estimate the average treatment effect using the following difference-in-difference model:[43]

$$\begin{aligned} P_{it} = \alpha_1 PlannedShut_{it} + \alpha_2 PlannedDowntime_{it} + \beta_1 UnplannedShut_{it} \\ + \beta_2 UnplannedDowntime_{it} + \gamma OutageAnyPlant_t + Time_t + Plant_i + \varepsilon_{it} \end{aligned} \tag{1}$$

where the sample is comprised of air pollution at the plant-day level 2014-2019. Plants include the 101 refinery plants existing in both EIA and Refinitiv datasets as discussed in Section 4.2. P_{it} is the average HCHO within 20km of refinery plant i on day t. $PlannedShut_{it}$ is an indicator that equals one if day t is the first day of a planned outage event for plant i and zero otherwise. $PlannedDowntime_{it}$ is an indicator that equals one if day t is the second to the last day of a planned outage event for plant i and zero otherwise. Similar dummies are coded for $UnplannedShut_{it}$ and $UnplannedDowntime_{it}$. While pollutant emission is close to zero when production is disrupted, there may be pollutant release on the first unplanned outage days. Separating the shutdown day and later days allows me to estimate the pollution response to the closing event and to the production suspension.[44]

In terms of other independent variables, $OutageAnyPlant_t$ indicates there is at least one plant under planned or unplanned outage on day t and is the same for all sample plants. I add plant fixed effects to remove any plant-specific time-invariant unobservables. I also add year, month, day of week fixed effects and quadratic time trends as additional controls to adjust for national time-based confounders. With time and plant fixed effects, the dummies $PlannedShut_{it}$, $PlannedDowntime_{it}$, $UnplannedShut_{it}$, and $UnplannedDowntime_{it}$ serve as the interaction terms in difference-in-difference to identify the impact of temporary outages on surrounding pollution, relative to pollution near plants without outages.

Coefficients α_2 and β_2 measure the impact of plants' normal operation on surrounding pollution and are expected to be negative. β_1 shows the impact of unplanned shutdowns on pollution. If β_1 is estimated positive, unplanned outage days result in excess pollutant release. The release

[43] Temporary outages take place multiple times and last for different durations. Only exploring responses within event windows requires me to stack multiple events for each plant and to choose the length of post-periods with a tradeoff between some plants' reopening and a short post-period.

[44] I am not using a distributed lag model for two reasons. First, I am interested in the average pollution decrease due to production suspension starting from the second day, so estimating one coefficient to measure the average effect is easy to interpret compared with assessing the individual impact on each day. Second, due to the different durations of each event, separating all days since the second day (i.e. estimating $UnplannedDay2_{it}$, $UnplannedDay3_{it}$, $UnplannedDay4_{it}$, etc.) would result in different sample sizes and precision of each estimate. As some events have long durations up to months, I will have a dummy for $UnplannedDayT_{it}^{+}$ eventually to capture all later days, so the distributed lag model will not help to capture individual effects on each day of the whole post-period.

events usually last for several hours, so the pollution increase should not be captured by β_2. Coefficient α_1 captures the impact of planned shutdowns on pollution. Planned outages are generally well designed and carefully implemented and should not induce pollutant release. α_1 should be negative or close to zero given the production suspension and the small estimation power.

The identifying assumption is that abnormal outages are assigned exogenously to any confounding environmental or socioeconomic features of the plants or their surrounding areas. This is true as abnormal events are unexpected, abrupt and emergent and usually put the plants off guard.[45,46] I use a similar design to assess the response of outcome variables, including tweets, crimes,[47] medical expenditure, foot traffic, and mortality.[48]

6 First stage results: air pollution response to outages

6.1 The effect of oil refining on air pollution

I use temporary outage schedules to identify the effect of plants' abnormal operations on surrounding air quality. Table S4.4 Panel A reports the estimated coefficients on unplanned outage dummies. HCHO increases by 0.6 units on the outage start day, equivalent to 7.2% relative to the mean HCHO levels. The abnormal shutdowns cause excess pollutant release and high emission events. As the events usually last for several hours, the effect becomes negative on the second to the last shutdown days, captured by estimates on $UnplannedDowntime$. In this period, the plant remains offline, and pollution is lower than the operating time. Estimates on $UnplannedDowntime$ indicate the normal operation leads to 0.22 units of HCHO increase in nearby areas, 2.8% relative to the average level.

[45] I conduct reduced form analysis due to the randomness and exogeneity assumption. Besides, unplanned outages are likely to satisfy exclusion restrictions for five reasons discussed in footnote 67.

[46] While I am primarily interested in the impact of abnormal outages, normal outages are also likely to take place exogenously, so β_2 indicates the causal impact of plants' normal operation on surrounding air quality. As shown in Figure S3.3, planned outages are correlated with low demand periods in the downstream product markets. In the U.S., fewer than 150 refinery plants serve the whole country's oil product use. Low demand in the product markets leads to low socioeconomic activities near product markets and low pollution levels there, but this happens far away from the production side and is unlikely correlated with other activities near refinery plants.

[47] Tweets were collected within buffers near plants, so tweet outcomes are at the plant-day level. When analyzing crimes and health responses, outcomes are at the county-day level, and treatment is assigned based on plants' counties. Sample only includes counties with at least one refinery plant and at least one observation in downstream outcome data. Those without plants or without any observations are dropped from the estimation.

[48] In my main specification of reduced form analysis, I use the same design as the first stage estimation. In Section 9, I show social media responses spread to geographically distant but online connected regions, which implies that ε_{it} is correlated with connected county j's pollution and $UnplannedShut_{jt}$. This does not violate my identifying assumption since both local and faraway pollution shocks are unexpected. However, controlling for non-local pollution shocks helps to explain the error term, improve the precision of estimates, and increase the overall model predicting power. I use an alternative specification in Section 9.6 to capture the inward spillovers in county i, namely the effect of local pollution conditional on non-local pollution shocks. Estimates on β_1 slightly decrease by 1.5%-3.5% and are not statistically different with this alternative specification. Compared with non-local pollution, local pollution contributes to observed offensive online activities to a much larger extent. Moreover, though social media and crime effects are related to online activities, health outcomes are biologically due to direct pollution exposure. Using equation (1) could make the reduced form analysis consistent across outcomes in Section 7 and 8. Therefore, I stick to the same design in my main analysis on downstream outcomes and provide additional discussion on inward spillovers in Section 9.6.

Another question of interest is the effect of plants' normal operation. Table S4.4 Panel B shows the effect of planned outages on surrounding HCHO levels. Estimates on *PlannedDowntime* are positive and significant, and remain stable with alternative fixed effects and time trends controlled. In the downtime period, HCHO within 20km of the closed plants is 0.5 units lower relative to the operating plants on the same day. The effect is equivalent to 5.6% relative to the mean HCHO levels.[49]

I put unplanned and planned outage schedules together and estimate equation (1). Results are reported in Table 2 Panel A. Estimates on *PlannedDowntime* and *UnplannedDowntime* are negative and precise, which suggests normal operations of refinery plants result in 0.2-0.45 units increase in surrounding VOCs. Compared with planned downtimes, unplanned downtimes lead to lower air quality improvements. This is likely due to fewer affected equipment and lower offline rates of unexpected failure than regular maintenance and security checks. Estimates on *UnplannedShut* are positive and significant, indicating the abnormal shutdown causes nearby HCHO to increase by 0.6 units. In contrast, estimates on *PlannedShut* are negative and imprecise. Planned shutdowns are well designed and prepared and do not generate pollutant leakage. Planned shutdown day 1 has a similar improvement on surrounding air quality as later days due to the production suspension.

Since planned shutdown day has no pollutant release, I combine *PlannedShut*, *PlannedDowntime* and *UnplannedDowntime* as one dummy *Downtime* and report results in Table 2 Panel B.[50] Results are similar to those in Table S4.4, which indicates planned and unplanned outage schedules are not highly correlated. Shutdown days of unplanned outages increase surrounding HCHO by 7.4%. Due to production pause, unplanned outages starting from day 2 and planned outages decrease HCHO by 4.3%. The regression controls include plant, day of week, year and month FEs in Column (1), and event year by month FEs in Column (2), thereby restricting comparisons to be within the same month. In Column (3) and (4), linear and quadratic time terms do not affect the main estimates, and do not improve the model fit captured by similar R^2. The negative estimates on *Days* indicate HCHO near plants improves over time.

Figure S3.5 shows event study figures of HCHO 10 days before to 10 days after abnormal outages. The regression controls include plant, year, state-month and day of week fixed effects. Figure 2 Panel A uses plants with capacity cover 200,000 bpd.[51] HCHO is not statistically different from zero on day -10 to -2, indicating zero pre-trends and the exogeneity of the abnormal outage schedule. HCHO increases on the unplanned outage start day and then decreases. The finding is consistent with the point estimates in Table 2. Negligible effects are found in the post-period due to short durations of unplanned events.

[49] Operating plants aside, I also study the effects of recently retired and reactivated refinery plants on surrounding air quality. Data, methods and results are summarized in Section S12.

[50] While my primary interest is abnormal operation, I still include *Downtime* on the right-hand side as normal downtime improves air quality and lowers P_{it} in my control group (non-abnormal outage plant-days). Estimate on *UnplannedShut* captures the poorer air quality due to abnormal outages compared with normal operation time which is the default condition in refinery plant zones.

[51] As discussed in Section S5, the effect of abnormal operations is mainly driven by large plants, so I focus on plants with capacity above 200,000 bpd when interpreting the event study figure.

Furthermore, I show abnormal outage is a strong instrument to predict surrounding pollution. Table S7.9 reports the estimated coefficients where HCHO is instrumented by unplanned outages. F-statistics are close to 30, indicating outage schedules could be used to evaluate the impact of pollution on downstream outcomes without weak instrument concerns.[52,53]

In summary, operating plants have strong and adverse effects on surrounding air quality. I find a 7.4% increase in HCHO near refinery plants on the abnormal shutdown day and a 4.3% decrease when the operation is suspended.[54] Abnormal operation of refineries generates 72% higher impacts on surrounding air quality than their normal operation. In addition, Section S5 reports heterogeneous effects of abnormal outages on pollution across outage timing and plant characteristics. I find outages on weekends, plants with larger capacity or using imported crude oil tend to have more severe pollution spikes when abnormal outages.

6.2 Robustness checks

There are several threats to my identification strategy, and I provide robustness checks to address these concerns. First, I include weather variables to control for weather conditions that may change air pollution. Weather variables include temperature (calculated as the simple average of daily maximum and minimum temperature), wind speed, and precipitation. The estimates in Table S6.1 remain stable with the inclusion of weather variables, which suggests that the changes in HCHO caused by temporary outages are not correlated with weather conditions.

Moreover, I explore whether abnormal outages could be predicted by weather conditions. I use the occurrence of unplanned outages as the dependent variable. In Table S6.2, controlling for national seasonality, annual trends, local temperature, precipitation, and wind patterns, I could not predict unplanned outages. Estimates on weather variables remain small and insignificant. To rule out the concern that unplanned outages are concentrated within a small number of plants, I run the regression with and without plant fixed effects. R^2 slightly increases from 0.001 to 0.006 using OLS, and pseudo R^2 stays the same using logit regression. This suggests plant fixed effects, as well as weather conditions, are not able to explain or predict abnormal outages. Similar findings are shown in Figure S6.1. Unplanned outages are well distributed across plants nationwide.

The third concern is also about the weather. Extreme weather events like hurricanes and storms could result in equipment failure and power outages, and in turn induce unplanned shut-

[52]More on the strong instrument, the recent literature on the first stage F-statistics suggests a more rigorous standard of 104.7 (Lee et al., 2021). The authors argue to adjust the critical value for t-ratio from 1.96 to 3.43 if the historical threshold for F-statistic of 10 (Staiger and Stock, 1997) is used. My F-statistic of 30.754 is lower than the threshold for reliable t-ratio-based inference. Alternatively, the Anderson-Rubin test has superior size properties compared to t-ratio-based inference even in the presence of weak instruments (Andrews et al., 2019; Lee et al., 2021). The 95% Anderson-Rubin confidence interval for Panel A Column (4) is [0.0257, 0.2484]. The Wald confidence interval is [0.0128, 0.2032]. Besides, the Anderson-Rubin Wald test has p-value of 0.0004 and F-statistic of 12.55. This confirms that estimates in Table S7.9 are highly significant and the instrument is strong.

[53]F-statistics are close to 30 using top ten plants in Table S7.9. Using all plants, F-statistics are above 10 in Table S7.16, S8.1, S8.9, and S8.11. This indicates a strong prediction power of outages on surrounding pollution, and there is no weak instrument concerns in later sections on downstream outcomes.

[54]The one-day pollution increase of 7.4% still generates detectible effects on downstream outcomes, discussed in Section 7 and 8. The statistical power of pollution effect is 0.69. Section 6.2 provides details on power calculation.

downs. These weather events could affect atmosphere conditions and socioeconomic activities that may directly or indirectly affect pollution levels. I therefore drop weather forecast-induced events for robustness checks. Table S6.3 shows similar positive estimates on $UnplannedShut$ and negative estimates on $PlannedDowntime$. The magnitude of pollution increase is remarkably similar to the main estimates, indicating that my findings are not driven by extreme weather events or by their induced behavioral responses. The standard errors are larger than those in Table 2 due to a smaller number of events.

The fourth check is on offline capacity. I include all outage events in the main result regardless of offline rates. If an outage has negligible effects on the operating condition, I expect lower effects on surrounding air quality. Therefore, I drop outages with low offline capacity below 50%, and 69.8% events are left. Results are displayed in Table S6.4. HCHO increase on the unplanned outage day is higher than that in Table 2, 0.6 vs. 1.2 units. The finding is consistent with the intuition that outage events with higher offline capacity should have more severe pollutant releases on the shutdown day and more substantial impacts on surrounding air quality.

The fifth check is on the sample construction. I use all the 101 plants existing in EIA and Refinitiv datasets as my sample. They have at least one planned outage or unplanned outage event 2014-2019. While planned outages are likely to happen every year due to maintenance, only 74 out of 101 plants encounter unplanned outages in the sample period. In the main specification, I define ever-treated plants on outage days as the treatment group, and other plant-days as the control group. Suppose the 27 never-treated plants with 0 unplanned events are quite different from the other plants, I may have an unbalanced sample and only capture the systematic difference between the never-treated and ever-treated plants. Therefore, I drop the 27 never-treated plants and re-estimate equation (1). Table S6.5 confirm the robustness of my results. The treatment effects of unplanned outages are not driven by the difference between never-treated and other plants.

The sixth concern lies in stacked treatment. 60 plants receive multiple abnormal outages during my study period. If the production process after the first outage changes or refinery operators learn from earlier pollutant releases, later outages may have different pollution patterns than the first treatment. Therefore, I use two other subsamples to compare pollution near treated plants when the first treatment occurs with pollution near never treated plants. In Table S6.6, I drop those 60 plants with multiple abnormal outages 2014-2019. Those experiencing only one outage have 2.1 units (26%) higher VOC when abnormal shutdowns compared with never-treated plants. In Table S6.7, I keep these 60 plants and drop days during and after the second treatment. Estimates on $UnplannedShut$ remain positive and significant, and show a 15% increase in surrounding VOC. The magnitude in both practices is larger than that in Table 2, which suggests the first abnormal outage in the sample period is more harmful than later ones. Learning by doing enables refinery operators to improve pollution controls in later abnormal events.

The seventh check is the decomposition of treatment effects. According to Goodman-Bacon

(2021), the estimate with two-way fixed effects is the weighted average of four components: comparing (i) early treated units to untreated units, (ii) late treated units to untreated units, (iii) early treated units to late treated units during the late group's pre-period, (iv) late treated units to early treated units during the early group's post-period. To explore the magnitude and weight of these components in the coefficient on $UnplannedShut$, I estimate equation (1) using `bacondecomp`. Results in Table S6.8 and Figure S6.2 show the largest weight (0.74) is put on 'T vs. Never treated', i.e. the double difference comparing plants with and without outages, during outages and normal operations. The other two components within the treated group have smaller weights. Note that component (iv) 'T vs. Already treated' does not exist in my setting, as my treatment duration is one day rather than a whole post-period. All these components have similar magnitudes to the main estimate.

In a similar vein, I check negative weight concerns based on de Chaisemartin and D'Haultfoeuille (2020) by using Stata package `twowayfeweights`. Under the common trends assumption, all estimates receive a positive weight and the sum of positive weights is 1. Under common trends and time-invariant groups' treatment effects, the sum of negative weights is -0.05, representing a very small contribution to the overall ATT estimate (the average treatment effect on treated). This test suggests there is not likely to be substantial bias in my estimated ATT due to negative weights.

I also test the statistical power of my estimation given the tradeoff between improving power and reducing confounding (Gelman and Carlin, 2014; Bagilet and Zabrocki, 2022). I use Stata command `retrodesign` (Linden, 2019) to compute the minimum sample size needed to provide precise estimates of the treatment effect. Parameter inputs include an effect size of 0.605, standard error of 0.246, and alpha equal to 0.05. Retrospective analysis shows a power of 0.6927, S-error of 0.000, and M-error of 1.2060.[55] My estimates are not likely to have lower power concerns or sign errors, which confirms the excess emissions of abnormal outages. Besides, the type M error of 1.2 denotes the factor by which the magnitude of an effect might be overestimated. The exaggeration ratio is close to 1, suggesting a similar value of the true and estimated effects.

The eighth concern is about restarting after outages. I only estimate pollution response during the outage duration and put all other days before and after the outage in the control group. If restarting plants results in excess pollutant emissions, I may underestimate the abnormal pollution spike due to the polluting control group. To address this concern, I add restarting dummies in the estimation. Table S6.9 reports positive but imprecise estimates on restarting dummies. Both planned and unplanned outages are associated with insignificant pollution increases on the restarting days. Estimates on outages are not affected by the inclusion of restarting dummies, indicating the robustness of my results.

Furthermore, air pollution is the average ambient level based on locations. One concern is regarding the overlapped 20km buffers. If plants are located close to each other and have

[55] Effect size and standard error are obtained from Table 2 Panel B Column (4). I use alternative alpha values as robustness checks. When an alpha value of 0.1 is used, power is 0.7935, S-error is 0.000, M-error is 1.1146; When alpha equals 0.01, power is 0.4552, S-error is 0.000, M-error is 1.3535.

different outage schedules, I may get biased estimates if pollution increases nonlinearly with multiple treatments. Therefore, I drop 33 refinery plants with overlapped 20km buffers and only use 68 isolated plants. Table S6.10 shows results are strongly robust with this alternative sample.

I also use ground-based benzene monitors to check temporary outages' effects on surrounding benzene. As is shown in Figure S3.2, only 46 plants have monitors in their counties, so I end up with a small sample size. In Table S6.11, benzene shows a 7.9% increase on abnormal outage days and a 3.3% decrease when normal outages. The magnitude is quite similar to the main results in Table 2. Furthermore, I use satellite aerosol optical depth (AOD) within 20km of each plant as dependent variable and replicate the first stage analysis. AOD data is obtained from two NASA satellites, TERRA, and AQUA with Moderate Resolution Imaging Spectroradiometer (MODIS). It is a daily product with a spatial resolution of 10km. AOD is well used in existing papers on atmospheric science and environmental regulations (e.g. Jin et al., 2019; Greenstone et al., 2020; Zou, 2021). Table S6.12 states AOD increases by 12.2% on the abnormal shutdown day. The magnitude is larger than that of the HCHO increase. In contrast, the magnitude of AOD increase due to refineries' normal operations is not serious enough to be detected by remotely sensed aerosol thickness.

As the analysis on downstream outcomes is mostly at the county level, I replicate my analysis using HCHO at the county-day level. To do so, I replace the dependent variable with the average HCHO in the county where each plant is located. I reach a similar conclusion in Table S6.14. This suggests that the production suspension of refinery plants improves HCHO in that county, while the unplanned outage days make the county's pollution worse.

To address the possibility that the results are not due to outage schedules but other confounding factors, I conduct a falsification analysis using placebo treatment counties. They are counties without refinery plants in the same states and should not correspond to any pollution changes in outage schedules. I find small and statistically insignificant estimates with opposite signs on outage dummies in Table S6.15.

Finally, I add the placebo counties to the sample in Table S6.14 to control for unobserved statewide conditions that may affect air pollution. I conduct a double difference analysis and report results in Table S6.16. Estimates on the single time terms are small and imprecise, while those on the interacted terms are similar to the main results.

7 Reduced form: local aggressive behaviors

First stage results in Section 6 indicate that abnormal refinery outage start days raise surrounding VOC levels, especially when the events take place in large plants, old plants, plants using imported crude oil, and on weekends. This pollution effect is more severe than refineries' normal operation identified using planned outage schedules. The temporary planned and unplanned outages provide natural experimental settings to study the effect of air pollution on downstream outcomes. In this section, I conduct a similar event study analysis and report estimated effects

on tweet content, expressed sentiment, and crimes. The results indicate that residents near polluters post more aggressive content, talk more about pollution, but have similar happiness levels on Twitter when air pollution goes up. Additionally, there are more person-related crimes and hate crimes against Black people.

7.1 Tweet effects

First, I use social media data to test the impact of air pollution on tweet content and expressed sentiment. As discussed in Section 4.5.1, five outcomes of interest include the average sentiment score, the number and the proportion of tweets with air pollution keywords, health keywords, offensive content, and racist content. Before assessing tweets, I replicate the first stage regression to ensure air pollution responses to outages still hold for the ten plants, similar to results using the complete set of plants. Table 3 Panel A reports the estimated coefficients. Unplanned outage day 1 increases surrounding HCHO by 5.3 units, 62.2% relative to the mean. The magnitude of pollution increase is larger than that in Table 2. This is likely due to the larger capacities, medium ages, and locations of the ten plants. As discussed in Section S5, heterogeneity tests show more pollution spikes when capacities are large and a nonlinear relationship between plant age and pollution severity. Besides, these ten are operating in states with more refinery plants including Texas, Louisiana, and Oklahoma where air quality is poor and even backsliding in recent years (Baurick et al., 2019; Colmer et al., 2020; ALA, 2022). These plants may receive less stringent environmental regulations, so their pollutant release during abnormal outages is more striking.[56]

Using a reduced form setting, I replace P_{it} with the five outcome variables in equation (1). In Table 3 Panel B, pollution-related tweets rise from 0.56 to 1.03 percentage points on the unplanned outage day. The increase is equivalent to 81.9% relative to the mean proportion. Pollution-related tweets capture discussion, complaints, and public awareness of worse environmental conditions. The environmental awareness still increases with surrounding worse air quality despite the missing regulation on excess emission in most states and the insufficient coverage of ambient monitors. Besides, the finding also serves as a validation of the first stage result: people talk more about pollution when the surrounding HCHO increases.

Unlike pollution-related tweets, no significant impacts are found on health-related tweets or expressed sentiment. While estimates on *UnplannedShut* are positive in Table 3 Panel C, suggesting health-related tweets also increase with pollution, they are not statistically different from zero and are not robust with alternative fixed effects. In terms of sentiment, estimates on *UnplannedShut* are small and imprecise in Panel D.[57] People's concern about health does

[56] The positive estimate on *UnplannedShut* in Table 3 Panel A has a mean shift compared with that in Table 2. While the larger magnitude is not driven by one plant, these ten have different magnitudes of increases. Specifically, three out of ten have below 50% but above 10% pollution increases during pollution spikes, while the other seven have over 50% below 100% increases. The average number of outages is 28.1 for these top ten plants, and is 8.9 for the entire sample of 101 plants.

[57] One concern on the expressed sentiment is that pollution may affect the tweeting sample. If people are too sad to tweet due to worse air quality, the remaining users that still tweet may be as happy as those when there is good air quality, and I thereby capture zero effects of pollution on sentiment. To address this concern, I check

not increase though they feel surrounding pollution is higher. Nor does pollution affect people's expressed happiness levels. This implies that people are not aware of health risks of polluting events.

Results in Panel E and F indicate significant increases in offensive and racist content on abnormal shutdown days.[58] The number of offensive tweets and racist tweets rise by 34 and 6 near each plant on each day, 24.4% and 30.1% relative to the average number.[59] Despite the same overall happiness levels captured by sentiment scores, people are more offensive and more likely to post racist content due to higher air pollution. These posts are primarily toxic, derogatory, and inflammatory, compared with generally unhappy tweets. My findings serve as warnings and potential policy tools to detect aggressive and mentally-unhealthy neighborhoods, as well as mechanisms for the relationship between air pollution and crimes that I will explore in Section 7.2.[60]

Figure 2 Panel B displays the event study figures of HCHO using the top ten plants with tweet outcomes. I find larger estimates on day 0 and flatter lines in the pre- and the post-periods. HCHO increases by about 5 units on the unplanned outage start days. Figure S3.8 shows the event study figures of tweet-related outcomes. In Panel A, there is a spike in pollution-related tweets on day 0 by 0.5 percentage points. In Panel B and C, the number of offensive tweets and racist tweets increases by 100 and 20 on day 0. Estimated coefficients are not statistically different from zero on day -10 to -2, and 1 to 10, confirming the exogeneity of abnormal outages.

My results confirm people are more offensive online on polluting days. So who or which groups are people offensive against? To answer this question, I further classify each offensive tweet using four binary classifiers: anti-government, xenophobic, sexual, and racist.[61,62] These labels are constructed separately but tend to be positively correlated. Table S7.7 reports the number of tweets classified as one in each subcategory. The most common group is sexual. 42% of all offensive tweets are offensive against different sex groups. Besides, people are more likely

whether the number of tweets decreases on abnormal outage days. Results in Table S7.4 show no change in tweet supply near refineries on unplanned outage days. Therefore, I conclude that expressed sentiment does not have sample selection problems.

[58] This is different from the null result in Panel D. As discussed in Almond and Du (2020), the interpretation of the sentiment score is not straightforward. Inflammatory and ironic tweets may be labeled as positive, but have negative emotional impacts on readers. Therefore, it is possible that toxic and offensive content in tweets still generates real-world crime impacts even with the same overall sentiment on social media.

[59] While most users take place only once in my collected tweets, I add user fixed effects to explore the mechanism of increased offensive and racist content and improve the model fitting. Results in Table S7.5 show adding user fixed effects absorbs the magnitude and precision of estimates. This implies that the effects on offensive tweets and racist tweets are mainly driven by the component of tweet authors rather than different content posted by the same authors.

[60] One concern arises if people are naturally aggressive when mentioning pollution and talk more about pollution on outage days as shown in Panel B, I only capture the offensive feeling in pollution-related tweets rather than general aggressive behaviors on Twitter. I drop all pollution-related tweets defined in Section 4.5.1 and re-estimate Panel E and F. Table S7.6 displays positive, significant and robust estimates on *UnplannedShut*. Besides, summary statistics on outcome variables (Y-mean and Y-sd) are very similar to those in Table 3. This indicates that tweets not related to pollution drive my observed more aggressive behaviors, and the potential concern above does not hold.

[61] The classifiers are available here: https://github.com/DanLeePy/Singapore-Comment-Filter. I use the random forest model with 500 tree parameters.

[62] The last subcategory racist is one if the tweet is both offensive and racist. It is a subset of racist tweets classified and described in Section 4.5.1.

to tweet with hostility towards other race groups than authorities or foreigners. The last row shows tweets classified as offensive but are not in any subcategories. They are offensive against people in the same group or normal items in their daily lives.

Subgroup results from estimating equation (1) are reported in Table S7.8. The insignificant estimate on *UnplannedShut* in Column (6) shows people are not aggressive against people in the same group. Positive and precise estimates in Column (2)-(5) indicate people tweet offensive content against those of different gender, race, nationality, or social classes. The increase in anti-government content is 16.1%. The increase in xenophobic, sexual, and racist content is 23.0%, 24.9%, and 22.6% respectively. The smaller increase in anti-authority tweets shows people are not only complaining about bad air or loose environmental regulations. Instead, people are generally more aggressive against other social groups.[63,64]

The finding of pollution-related tweets implies the increased environmental awareness on abnormal outage days. This is consistent with existing evidence on people's adaptation to environmental stressors (e.g. Deschenes and Greenstone, 2011; Moretti and Neidell, 2011). As excess pollutant release usually lasts for several hours, one may ask whether the awareness effect and the psychological effect are also short-lasting. I therefore use pollution tweets, offensive tweets, and racist tweets on lead day 1 and 2 as outcomes, and those on lag day 1 as a placebo test. Table S7.10 Panel A shows estimates on *UnplannedShut* are still large and positive but not statistically significant. Consistent with the one-day pollution change, people only talk about pollution when it is polluting outside, and stop discussing it the next day. The sign of the estimates flips in Panel B and Panel C, indicating zero effects on the lead day 2 and lag day 1. Similarly, impacts of outage-induced air pollution on offensive and racist content are not significant on lead and lag days. The online aggression responses are consistent with medical evidence that air pollution induces immediate same-day biological and behavioral changes (Musi et al., 1994). Despite the short-lasting effects, my finding serves as a signal for adaptation in response to poor air quality, and underscores future research on pollution and adaptation with better real-world data on measures of adaptation behaviors.

7.2 Crime effects

The second outcome I explore is crime. Table 4 summarizes estimated coefficients on the number of hate crime events against Black people and the number of Black victims. In Column (1), I add county, year, month, day of week fixed effects. I conduct a within-month comparison by adding year-month fixed effects in Column (2), and point estimates on *UnplannedShut* are

[63] I use a reduced form design to assess tweet responses on unplanned outage days given the exogeneity assumption. I also implement an instrument variable (IV) regression to check the magnitude of my results. Table S7.9 reports the estimated coefficients where HCHO is instrumented by unplanned outages. As HCHO increases by 1 unit, the proportion of pollution-related tweets increases by 0.1 percentage points, while the proportion of health-related tweets and sentiment scores are not significantly affected. Besides, the number of offensive tweets and racist tweets increases by 15 and 2.5 at the plant-day level, 10.9% and 13.5% relative to the average numbers.

[64] To get the causal effect of pollution on tweets, I also need the exclusion restriction assumption. It is difficult to confirm that unplanned outages only affect tweets through ambient pollution. Therefore, I don't attempt to interpret the estimate coefficients and focus on the reduced form results when discussing downstream outcomes.

remarkably stable. On abnormal outage days, the number of hate crime events increases by 0.02, 54.8% relative to the mean. There are 0.03 more Black victims associated with hate crime events, 78.1% relative to the average levels. Despite the small point estimates, the magnitude of effects is substantial compared with the average crime incidence. The impact on victims is more striking than that on event occurrences, which suggests pollution increases both the occurrence and the severity of hate crimes against Black people. I add linear and quadratic time terms in Column (3) and (4). The model fits do not get improved, indicating hate crime could not be explained by time trends.

Unlike the offending towards Black people, abnormal outages of refineries have no impact on hate crime against the white, Asian or Hispanic group, as is shown in Table S7.12 Panel A to C. Estimates on *UnplannedShut* keep small and imprecise in Column (1)-(4). The findings show a disproportionately adverse effect on Black people when pollution increases. This is likely due to the larger proportion of Black population groups living near refineries. As shown in Table S8.7, the proportion of Black people living within 25km of refinery plants is 11-21%. The proportion decreases to 7-13% more than 100km away from refineries. In contrast, only 58-80% of residents living close to refineries are white, while the proportion of white people 100km away from refineries is 76-89%. If air pollution increases the most near refineries as hotspots than elsewhere in the same county, I expect a more substantial effect on hate crimes near refineries. Given a different race proportion, the black people concentrated near refineries face higher risks of being offended.

Apart from hate crimes, I find positive and significant impacts of unexpected shutdown schedules on other crimes and victims. As shown in Table 5, there are 5 more crimes and 7 more victims on the abnormal shutdown day, 12.1% and 13.1% of the average levels. Planned outages and the second to the last day of unplanned outages are associated with insignificantly fewer crimes. The magnitude of non-hate crime increase is smaller than that of hate crimes, implying a more considerable impact on physical violence and racial bias. Besides, R^2 is over 0.5 in Table 5, larger than that in Table 4, 0.15. Compared with non-hate crimes, hate crimes could hardly be explained by outage schedules, county and time fixed effects, and time trends added in the estimation.

Why do hate crimes respond more to pollution than other crimes? I propose three potential reasons. First, hate crimes are more social and person-related. Compared with other crimes, hate crimes are more closely related to social cohesion, trust, control, and disorganization (Lyons, 2007). Hate crimes move together with adverse changes in social interactions. Second, hate crimes capture negative attitudes towards out-group members, which matters less in other crimes. As hypothesized in Section 3, air pollution affects expressed aggression, captured as more severe out-group negativity rather than attacks towards in-group members. Third, the expected penalty for hate crimes is lower. While the punishment of hate crimes is more severe than non-hate crimes in general (Mellgren et al., 2017), the underreporting rate is higher (Pezzella et al., 2019), so the low expected punishment increases the incentives of committing

hate crimes.[65]

Compared with an air pollution increase of 7.4%, a crime increase of 12.1% has a larger magnitude, probably due to three reasons. First, I measure air pollution using 20km buffers near plants due to satellite product resolution. Refinery plants' emissions tend to concentrate near plants, and the nearby air pollution increase could be more striking than 7.4%. In Table 6, I separately estimate benzene increases using monitor data. To do so, I merge plants with surrounding monitors using distance cutoffs between 1km and 35km. Estimates on *UnplannedShut* gradually decrease as distance increases, suggesting pollution increases are more striking near plants. The 7.4% increase in satellite VOC within 20km buffers is an average of 72.7% increase within 1km buffers and moderate increases in surrounding farther areas. Therefore, the 12.1% increase in crimes is smaller than air pollution increases in plants' close atmosphere. Note that Y-mean and estimates on *Downtime* also decrease with distances, which indicates generally higher pollution near plants than farther surrounding areas.

Second, air pollution is multi-dimensional. Given a notable pollution shock with substantial pollutants released, I expect other pollutants also change with outage schedules. I focus on VOC changes in my first stage analysis as VOC is the immediate and major pollutant from refineries. I also find AOD increases by 12.2% in Table S6.12, and I expect other pollutants are also emitted together, which together results in bad air quality and drives the downstream impacts. Using a limited number of monitor reports in plants' counties, I find SO_2, NO_2, and CO insignificantly increase, and the magnitude ranges between 6.5%-9.7%, shown in Table S6.13. While I could not disentangle specific pollutants impacts, I attribute my downstream impacts to air pollution changes, not only VOC changes.

Third, the daily variation of air pollution is much smaller than the daily variation of crime events. In my sample refinery counties, the standard deviation in air pollution is 1.2% on average, much smaller than a 7% increase on the shutdown day. In contrast, the standard deviation in crime is 51%, so a 12% increase in crime is not large relative to its natural variation.[66]

Another question is why the estimated elasticity of tweets to hate crimes is greater than 1. I provide two explanations. First, a 24-30% change in tweets and a 59% change in crimes have different base values. The total number of tweets is much larger than the total number of crimes. Second, offensive or racist tweets do not always generate crime. Among these aggressive tweets, if technically doable, I could separate them into two groups: those that are signals for actual crimes and those that are not. I expect different responses in these two, with more striking effects on the former. Similarly, another way around, I could separate real-world aggressive behaviors into non-crimes and crimes. After people become aggressive, in the real world, multiple types of grumpy behaviors change, like aggressive driving, impulse buying, and saying swear words.

[65] If air pollution is perceived as a threat, local residents may act aggressively as a protective response, so they attack others to fight for avoidance shelters and commit more person-related crimes. This is not true given my non-local findings in Section 9.3. Physical violence still increases even if remote areas have no change in pollution and environmental stressors are no longer threats for these remote residents.

[66] In a similar idea, I also find the cross-sectional variation in crime is about 180%, and my finding on crime effect is below one tenth relative to cross-sectional variation.

Crime is only a subset of them and has a larger effect than non-crime grumpy behaviors.

Among non-hate crimes, I test subtype responses in Table S7.15. Homicides and assaults rise by over 16% on the first day of unplanned shutdowns. In contrast, estimates on drug and narcotic offenses are small or imprecise, indicating minor impacts of air pollution. My findings show that person-related crimes increase more than property- or society-related crimes as surrounding air quality gets worse. Person-related crimes are more associated with impulse control, as discussed in Section 10.1. In contrast, property- and society-related crimes are usually prepared ahead of time and mainly driven by income effects that are less likely to respond to refineries' abnormal outages.[67,68] Similar findings are documented in Heilmann et al. (2021) who show violent, domestic, and intimate crimes have more significant increases than property crimes on hot days, another important environmental stressor affecting aggressive behaviors.

Table S7.14 reports the impact of abnormal outage days on hate crime on lead and lag days. Similar to tweet-related outcomes, the effect on hate crime is also short-lasting. Estimates on *UnplannedShut* are not statistically significant in all these panels. The crime level returns to normal one day after the unplanned outage days. Results in Table S7.17 display similar short-lasting impacts on non-hate crimes. The prompt and same-day effects are similar to existing lab experiments and observational evidence on air pollution and physical aggressive behaviors (Musi et al., 1994; Burkhardt et al., 2019).

[67] The impact of refinery outages on downstream outcomes is not likely to result from economic activities for five reasons. First, outages are not disclosed to the public. Nor do people experience any effects in their daily lives. Changes in health and crime outcomes are solely driven by breathing bad air. Second, regarding economic activities and energy markets, refineries serve as upstream facilities that provide inputs to downstream markets. Effects in downstream markets are not that prompt, and are geographically far away from refinery plants. Third, refinery workers are the only group that know and experience the event, but they are a small fraction of population and are busy inside plants to get them back to normal. It is not likely that workers are free from work, outdoor, and commit crimes instead. See footnote 68 for more discussion on workers. Fourth, estimation results in Section 8.1 and 8.2 show there is no effect on overall outdoor activities, police governance, and placebo product purchase, which confirms that economic activities are not changed due to refinery outages. Fifth, as a proxy for daily income, I use expenditure on products whose income elasticity of demand is high or negative as outcome variables. More purchase of luxuries measures higher income, while more purchase of inferior goods measures lower income. I use electronics, video games, and automobile-related products for luxuries, and instant coffee and tea, prepared and frozen food as inferior products. Results in Table S8.3 show abnormal outages have no detectable effects on these proxies for daily income. Those with high income elasticity and negative income elasticity have similar point estimates on *UnplannedShut*, and results are statistically insignificant, confirming income channel is not the driver of the outage impacts.

[68] According to the Bureau of Labor Statistics, there are 169,195 petroleum refining employees in the U.S. in 2019. Among them, 32,138 are working in business and management positions. On average, 1,253 employees are hired by each plant, 238 management and 1,015 non-management. Refinery plants have continuous process operations, operating 24 hours per day. Workers are arranged in three-cycle shifts (Taylor, 1967; Ljoså and Lau, 2009). That said, in each hour, one-third of workers out of the total are on patrol, ignoring balance days and holidays. Therefore, 338-576 employees, 0.33-0.56% population in each refinery county are directly affected by outages, a small fraction of the local population. The fraction is also small compared with the tweet author set. There are 25 million collected tweets and over half of the authors write only one or two tweets between 2014-2019. Furthermore, I find no change in tweets with job-related keywords, so Twitter users are not concerned about job security or unemployment due to unexpected outages. Job-related keywords and estimation results are reported in Table S7.11.

8 Reduced form: local health outcomes

8.1 Medical expenditure effects

I evaluate changes in consumption expenditure in response to abnormal outages. I use medical-related products as a proxy for health conditions and hypothesize that pollution worsens surrounding residents' health. Table 7 presents results on cough remedies, sinus remedies, and breathing aids. All these three products are related to respiratory systems and serve as prompt aids for acute failures. In Column (1), expenditure on cough remedies increases by \$0.65 at the county-day level, 58.6% relative to the mean.[69] In Column (2) and (3), consumer spending on sinus remedies and breathing aids increases by \$0.07 and \$0.03 on the treated days, 38.0% and 96.8% relative to the average expenditure. The magnitude of the effect (38%-97% increase in respiratory aids) is remarkably large. Abnormal outage events are highly polluting and short-lasting, so they are more likely to induce acute health responses compared with continuous and mild pollution. Despite the transient effect of one day, health expenditures confirm the substantial burden faced by surrounding residents and underscores the severity of the pollution problem.

Medical expenditure data is recorded with household demographic information. This enables me to test heterogeneous effects across race groups. The hypothesis is that black households live near refineries and have extra higher health burdens than the white group in the same county. I add *Black* dummy and its interaction with *UnplannedShut* in equation (1). Shown in Table S8.6, estimates on the single term *Black* capture the baseline difference in medical expenditure in Black surveyed households relative to white surveyed households. Black households spend 145% less on cough remedies, 6.4% and 150% less on sinus remedies and breathing aids, probably due to lower disposable incomes. Estimates on *UnplannedShut* remain robust with the inclusion of race terms. Coefficients on the interaction terms are negative and insignificant, which implies that Black people are not more adversely affected by emission events. In other words, both Black and white people face similar short-term health burdens with the same pollution increase.

Despite similar per unit effects, we still have environmental injustice of refinery pollution, as minority groups live close to refineries. Each household has the same health burden when excess emission events, but the total health burden of Black people is higher than that of other groups. In Table S8.7, I report the race composition by distance to refineries in each 25km interval. I use data from the American Community Survey (ACS) and the Nielsen surveyed households.[70] The proportion of Black people decreases as the sample moves farther away from refineries, while the proportion of white people increases. This indicates the Black group disproportionately suffers exposure to refinery pollution and related health effects.

Abnormal outages have no impact on placebo products' consumption. In Table S8.2, esti-

[69] To collapse data from the expenditure level to the county-day level, I calculate the sum of expenditure of all surveyed households in each county on each day. The absolute value of expenditure is small due to the small number of households in each county, so I focus on the proportion change relative to the mean when interpreting the results.

[70] This practice also confirms the Nielsen consumer panel is representative of the U.S. population.

mates remain small and imprecise with consumption expenditure on eyeglass accessories, tooth and gum analgesics, or insoles as dependent variables. These products are not related to air pollution or health and should not respond to pollution events.

In terms of expenditure on lead and lag days, Table S8.4 shows similar zero effects of abnormal shutdown days. None of these three purchases responds to pollution on the other days. The findings on medical expenditure are compatible with those on tweet-related outcomes and crimes. Pollution induces immediate effects on the same day.

I implement a robustness check only using households with complete data every year 2014-2019. The sample selection concern exists if households affected by air pollution choose not to respond to the Nielsen survey that year. There are 31.3K households responding to the survey every year 2014-2019, and 109.2K households responding at least one year. The latter one comprises the sample for the main analysis in Table 7. Using the 31.3K households with complete data, the effect of unplanned outages on medical expenditure still holds, and the magnitude gets more prominent, as is shown in Table S8.5. Household expenditure on cough remedies and sinus remedies increases by \$0.6 and \$0.1 dollars at the county-day level when abnormal outages, 88.6% and 73.6% relative to the mean. Expenditure on breathing aids insignificantly increases by \$0.01 dollars or 56.3%. The results are robust using the strict sample without sample selection concerns.

8.2 Foot traffic effects

Another health outcome of interest is hospital visits, captured by people's movement patterns. As the time range of foot traffic data is 2018-2019, I replicate the first stage analysis only using two years' sample. Results in Table S8.8 show similar positive estimates on $UnplannedShut$, and negative estimates on $PlannedShut$, $PlannedDowntime$, and $UnplannedDowntime$. The magnitude of pollution increase on abnormal shutdown days is larger than that using six years' data, 1.0 vs. 0.6 units HCHO increase. The precision of estimates is lower due to the small sample size and fewer outage treatments. In general, the first stage results hold in a shorter time period.

I use visits to health-related POIs, visits to outdoor leisure-related POIs, and overall county busyness as outcomes. In Table 8 Column (1), abnormal outage days have no significant impacts on counties' overall busyness, captured by the number of active devices. In Column (2), the number of visits to amusement parks and recreational camps is also unaffected. The findings show emission events do not affect outdoor leisure and overall movement activities. People don't have sufficient information on outage-induced pollution and keep their outdoor trips. In Column (3), abnormal shutdown days increase the number of visits to hospitals by 249K, 20.8% relative to the mean. The results are consistent with those in Section 8.1 that air pollution affects health and increases medical-related activities.

Effects on people's movement on lead and lag days are shown in Table S8.10. In Column (1) and (2), I find small and insignificant estimates on $UnplannedShut$. Estimates on $Downtime$

are negative and significant in all these panels and in Table 8. In Column (3), the estimate on *UnplannedShut* on lead day 1 is positive and large, and that on lead day 2 becomes smaller. Despite the same signs, the effect is not significant on both lead days, suggesting only the same-day hospital visit increases with polluting events. The placebo estimate on lag day 1 is small, negative and imprecise.

8.3 Mortality effects

The last local outcome I analyze is mortality. I use county-day level mortality and mortality rate to assess the effect of refinery pollution. Table 9 summaries the reduced form results. In Panel A, on the abnormal shutdown day, the number of deaths in the county where plants are located increases by 0.03, 3.3% relative to the average. In Column (1)-(4), estimates on *UnplannedShut* remain positive but imprecise, implying the mortality increase is not statistically different from zero. Besides, planned outages and the second to the last unplanned outage days lead to a decrease in mortality by 0.05, 5.5% relative to the mean. Again, estimates remain statistically insignificant. Similar patterns are observed in Panel B. The proportion of deceased beneficiaries does not respond to refinery outage schedules.

My results are likely to suffer from low power concerns, given the insufficient coverage of the Medicare program and the 5% beneficiary data. In addition, there are two potential reasons for the little response in mortality when refineries experience unexpected emission shocks and surrounding air quality gets worse. First, Medicare beneficiaries may not live close to refineries, so my county-level analysis may not capture mortality responses for those near refineries. Second, these beneficiaries tend to have insurance coverage and available medical resources. Even if outage-induced emissions lead to air pollution spikes, beneficiaries do not encounter extreme health loss. They compensate for their health reduction by increasing their consumption of medical-related products, as discussed in Section 8.1. As a result, there is no significant increase in deaths among beneficiaries.

9 Spillover: non-local aggressive behaviors

Personal networks are no longer restricted by geography and physical space with the help of social media and online connections. Much like in the offline world, emotions on social media can be transferred from one person to another. Existing studies document that positive and negative moods are correlated in networks (Fowler and Christakis, 2008; Rosenquist et al., 2011). In an experiment with Facebook users, after researchers manually deleted positive content, other Facebook users produce fewer positive posts and more negative ones (Kramer et al., 2014). In the context of pollution impacts, local air pollution leads to more offensive posts on social media at the local scale. In this section, I test whether this emotion contagion spreads to geographically distant areas through social networks. Compared with regions loosely connected with polluters, distant but closely connected regions are hypothesized to witness more offensive content on social

media. I also assess connected areas' crime responses to faraway pollution shocks, given social media's contribution to inflaming conflicts.

9.1 Data and empirical strategy

The unexpected outages lead to more negative tweets near refineries. I exploit the exogenous variation in emotional contagion through the cross-sectional variation in social connectedness with other geographically distant areas. The hypothesis is that when an individual sees Twitter connections posting offensive tweets, the possibility of him being offensive becomes higher. In contrast, in areas where residents are not closely connected to those near refineries, a user has fewer Twitter connections on her home page writing hostile content, so her emotion and online activities are less affected.

I obtain the cross-sectional social connectedness index (SCI) at the county-county level, August 2018 version.[71] SCI is developed by (Bailey et al., 2018) by aggregating anonymized information from friendship links between all Facebook users. Facebook is the most popular social network with more than 2.7 billion active global users and 231 million U.S. users, 70% of the U.S. population. SCI measures the probability that two individuals across two locations are friends with each other on Facebook, and is specified as:[72]

$$SCI_{ij} = \frac{Facebook\ connections_{ij}}{Facebook\ users_i \times Facebook\ users_j}$$

I use Facebook SCI to test emotion contagion on Twitter. According to (Bailey et al., 2018) and (Kuchler et al., 2020), Facebook users in the U.S. typically use this platform to interact with friends beyond the Facebook platform. The measure of social connectedness, although derived from Facebook friendship links, captures the exchange of information among general online networks and the strength of connections between U.S. counties.

I validate the reliability of the SCI measure in my context by testing the correlation of SCI and Twitter following relationships. To do so, I use tweets written near ten refinery plants analyzed in Section 7.1. There are 77,122 bad authors that wrote at least one offensive or racist tweet 2014-2019. These bad authors have 969,961 followers. Among all followers, 161,501 (16.7%) have county information in their profile locations. I use the following link-level data and calculate the number of followers at the county pair level between bad authors' county and followers' county. Then I calculate the correlation of follower count and Facebook SCI between these 10 plants' counties and all other US counties.

[71] I obtained the data through a non-disclosure data-sharing agreement. A newer version of SCI, October 2021, is open for download (https://data.humdata.org/dataset/social-connectedness-index?fbclid=IwAR3V5LNhpIEIPC4N8ctJ_fpsd49hiGh-yuEpR1MrKFrEVllJsaUPpWRY-Vw). Given my sample period, 2014-2019, I use the August 2018 version in case the COVID affects friendship between counties.

[72] SCI data is well used in the economics literature to study online connectedness and its impacts. Existing studies find users with online networks affect people's decision to buy or rent houses (Bailey et al., 2018), to adopt new cellphone products (Bailey et al., 2019), to participate in flood insurance (Hu, 2020), and to follow the COVID control policies (Fritz and Kauermann, 2020; Holtz et al., 2020b). Social networks also increase technological diffusion (Diemer and Regan, 2020), returns for acquirers (Nguyen et al., 2021), inter-regional bank lending (Rehbein and Rother, 2020), and trust in mutual political parties (Bailey et al., 2020).

$$SCI_{ij} = \#Followers_{ij} + \phi_i + \varepsilon_{ij}$$

where SCI_{ij} is the Facebook SCI between refinery county i (i.e., bad author county, followee county) and followers' county j. $Followers_{ij}$ is the number of followers in county j that are following bad author in county i. Results in Table S9.1 suggest that these two matrices based on Facebook and Twitter are positively and significantly correlated. This practice confirms that online social media platforms are mutually correlated, and Facebook SCI could be interpreted as the strength of general online networks across regions. Results using this follower-based Twitter SCI are reported in Section 9.4.

Apart from constructing my own version of Twitter connectivity matrix, I validate SCI data by conducting an individual-user level analysis in Section 9.2. Since I observe each tweet author's Twitter handle, I collect follower handles and follower tweets and confirm that follower tweets are affected by followees' local shocks. The user-level analysis shows a similar result of online emotion contagion as the county-pair level results using SCI.

I use an event study analysis to test the impact of outage schedules on distant tweets and crimes by estimating the following equation:

$$\begin{aligned} Y_{jt} = \beta_1 UnplannedShut_{it} \times Connected_{ij} + \gamma OutageAnyPlant_t \\ + Time_t + County_j + \varepsilon_{jt} \end{aligned} \tag{2}$$

where Y_{jt} denotes tweet or crime outcomes in recipient county j on day t. $UnplannedShut_{it}$ is an indicator that equals one if refinery i starts an unexpected outage on day t and zero otherwise, same coding as the local event dummy in equation (1).[73] $Connected_{ij}$ is a continuous connectedness measure derived from SCI data. Here I use an absolute measure and a relative measure. If recipient county j is socially distant from all other counties, but relatively more connected with the polluter's county i, I expect a large value in the relative measure but a small value in the absolute measure:

$$Connected_{ij} = ln(SCI_{ij})$$

$$Connected_{ij} = \frac{SCI_{ij} * Pop_j}{\sum_K (SCI_{ik} * Pop_k)}$$

$UnplannedShut_{it} \times Connected_{ij}$ is the treatment variable of chef interest.[74] To aggregate R refineries, I use the maximum treatment of a weighted average where w_r is either equal-weighting

[73] In equation (2), connected plants' production suspension period is not controlled due to its insignificant effects on local tweets that are treatment in remote outcomes. I add $Downtime_{it} \times Connected_{ij}$ as robustness checks. Results in Table S9.23 and S9.24 show estimates on β_1 and R^2 are not affected by the newly added control. Estimates on $Downtime_{it} \times Connected_{ij}$ stay insignificant in all panels.

[74] I separately estimate effects on connected and unconnected counties using the top and bottom quartile. Results in Table S9.27 and S9.28 show positive, significant and similar estimates on β_1 using top quartile counties, while imprecise estimates using loosely connected ones. Therefore, my control group is not significantly affected by remote pollution spikes, and the main specification does not suffer downward bias.

or population-weighting:

$$UnplannedShut_{it} \times Connected_{ij} = \max_{R} UnplannedShut_{rt} \times Connected_{rj}$$

$$UnplannedShut_{it} \times Connected_{ij} = \sum_{R} w_r \times UnplannedShut_{rt} \times Connected_{rj}$$

On other right-hand side variables, $OutageAnyPlant_t$ is the same for all sample counties and indicates that there is at least one plant experiencing abnormal shutdowns on day t. Time controls include year, month, day of week fixed effects and quadratic time trends. I also add county fixed effects to control for baseline county-level differences in tweeting and crime activities.

Coefficient β_1 captures the causal effect of local pollution shocks on geographically distant outcomes through social connections. The identifying assumption is that geographically distant areas are affected by pollution shocks from refineries only through the channel of online networks. The information of pollution spikes due to outages is not disclosed to local residents near refineries, let alone distant residents. So other areas should not generate any information about the actual outage. Other possible threats to identification, including pollution dispersion, physical movement and oil product market, are discussed in Section 9.5. I only use counties that are 150km away from any refinery plant. Figure S3.16 shows counties in the sample that are closely or loosely connected with refinery counties. Alternative cutoff distances are used for robustness checks.

9.2 Results on tweets

Table 10 displays estimation results with the same five tweet outcomes as dependent variables.[75] In Column (1)-(3), sentiment, complaints about pollution, and discussion about health do not respond to connected counties' pollution shocks. Pollution-related tweets have small Y-mean and negative point estimate despite low precision. Compared with residents in the refinery zone, people in remote counties in general do not complain much about air quality. Positive estimates on the interaction term in Column (4) and (5) conclude that having online connections post aggressive content increases the likelihood of being aggressive oneself. Compared with other distant counties, those closely connected with refineries are more affected by pollution shocks. On abnormal shutdown days, the number of offensive and racist tweets increases by 3.36% and 1.43%, as connectedness increases by 1 unit (2.8 standard deviations, 48% relative to refinery counties' within-county connectedness). Subgroup results are shown in Table S9.10. On polluting days, connected areas see 3.24%, 3.40%, 3.43%, and 3.50% increases in anti-government, xenophobic, sexual, and racist tweets, respectively. The increase is comparable across subcategories. Compared with results in Table 3 (24.4% and 30.1% increases in offensive and racist

[75] Given the computation time and API quota constraints, I only collect geo-coded tweets for five minutes on each day to calculate the tweet outcomes. Figure S3.17 shows the diurnal cycle of tweeting activities near the top ten refinery plants. The number of tweets is small during the local wee hours, starts to increase around 7am, and is even higher after 6pm when most people leave work. I collect tweets during 21:10-21:11, 21:20-21:21, 21:30-21:31, 21:40-21:41, and 21:50-21:51. They are busy tweeting times and late enough to detect tweet reading effects on the same day.

tweets), the spillover effects to connected distant areas are about one seventh of the effects near polluters.

As robustness checks, Table S9.11 reports results using relative measure to code $Connected_{ij}$ and alternative aggregation methods. In Panel A, counties behave similarly in response to pollution shocks when using the relative measure of connectedness. It is the likelihood of being friends that drives the online emotion spread. In other words, areas that are generally not close with others but relatively more close with refineries are not affected, while regions connected with both refineries and other counties receive the spillover effects. In Panel B and C, estimates in Column (4) and (5) are similar to those in the main specification. Besides, I use alternative distant cutoffs to define geographically distant counties, and the results remain strongly robust in Table S9.12. In addition, collected tweets from API include retweeted tweets with comments (0.02%) and original tweets (99.8%), and do not include pure retweets without comments. Despite a low proportion, I drop retweeted tweets in case they include aggressive content originally tweeted by users near polluters. Results only using original tweets without other users' text are reported in Table S9.13. They are very similar to those using all collected tweets in Table 10.

Apart from a county-aggregated analysis, I perform a user-level analysis since these aggressive authors' Twitter handles and their follower lists are available. I focus on one refinery and its surrounding 'bad' authors who posted at least one offensive tweet or racist tweet and collect their followers' tweets 2014-2019.[76,77] There are 1,109 bad authors who posted 89,291 offensive and 16,942 racist tweets. Of these followees, 2,683 followers posted 2,810,392 tweets 2014-2019. Among them, 171,669 are classified as offensive, and 8,757 are racist. The hypothesis is that reading followees' aggressive posts increases the likelihood of followers' being aggressive. I conduct a correlation test and a reduced form analysis using the following equations:

$$Offensive_{ijt} = \alpha_1 Offensive_{it} + Follower_j + Time_t + \varepsilon_{ijt} \tag{3}$$

$$Offensive_{ijt} = \beta_1 UnplannedShut_{it} + \gamma OutageAnyPlant_t + Follower_j + Time_t + \varepsilon_{ijt} \tag{4}$$

where the sample in equation (3) includes tweets at the bad author-follower-day level. $Offensive_{ijt}$ is a dummy that equals one if bad author i's follower j posted at least one offensive tweet on day t. I also use the number of offensive tweets to test the impact on the intensive margin. On the right hand side, $Offensive_{it}$ is an indicator that bad author i posted at least one offensive tweet on day t. I add follower j fixed effects to control for tweet author-specific time-static features. Coefficient α_1 captures the correlation of bad authors' being offensive and their followers'

[76] I use ExxonMobil Baytown refinery in the user-level analysis. Detailed information of this plant could be found in Table S7.1 Column (1). Collecting 2,810,392 tweets written by bad authors' followers requires three months' API quota. Given the quota constraint, I only focus on one plant rather than all plants for this user-level analysis.

[77] The number of followers of my sample bad accounts near ExxonMobil Baytown refinery ranges between 2 and 100, with an average of 85.6. According to Patrikarakos (2017), the average Twitter user has 208 followers. Therefore, my sample authors that wrote at least one offensive or racist tweet have fewer followers than representative users. The results of more offensive tweets in connected areas are not likely driven by several giant accounts, but driven by a large number of ordinary unpopular accounts.

probability of being offensive.

Results in Table S9.2 show followees' and followers' aggressive behaviors have significant and positive associations. When bad authors post an offensive tweet, followers are more likely to be offensive by 1.2 percentage points or 80% compared with their own tweets on other days. For racist content, the estimation precision increases, and the racist likelihood goes up by 0.1 percentage points or 76.4%. Unlike the extensive margin, the number of aggressive tweets shows positive but statistically insignificant responses to followees' posts.

In the reduced form analysis, the sample in equation (4) is also at the bad author-follower-day level and only includes bad authors near refineries. I use a similar design as equation (2) to assess the impact of refinery pollution shock near bad author i on follower j's tweets, captured by coefficient β_1. Results displayed in Table S9.3 suggest abnormal outages also make followers more aggressive. The magnitude of the increase is 16.1% and 20.7% for offensive and racist content, 66% of the local effect and 4.9 times the county-aggregated non-local effect. Compared with county-level spillover results in Table 10, effects are larger for three reasons. First, authors in Table S9.3 follow someone near refineries, while authors in Table 10 include all accounts in distant recipient counties. The latter may not follow any accounts or any accounts near polluters. Second, authors in Table S9.3 follow offensive tweet authors near refineries. If reading offensive tweets is the treatment, they have a stronger first stage than authors in Table 10.[78] Third, mechanically, people following bad authors are not representative of the general public or all social media users. They tend to be a more aggressive subgroup, so they may respond differently on pollution shock days. To sum up, the user-level analysis confirms the online emotion spread identified in county-aggregated results. While the user-level effect is important, there are substantial heterogeneities across user characteristics and county baseline conditions. The aggregate results may be more informative from the policy perspective to indicate geographical spillovers and quantify non-local effects of pollution.[79]

9.3 Results on crimes

In Table 11, Column (1) shows the number of hate crimes increases by 1.5% on the outage day in connected counties relative to unconnected ones. The increase is statistically insignificant. Similar patterns are found in subgroup results. As connectedness increases by 1 unit, the number of hate crimes against Black people insignificantly increases by 32%. The estimate is negative in Column (3), suggesting Asians are not more severely attacked with more offensive content online. Despite the low precision, the magnitude of estimates in Panel B is larger, which implies results on the following day are more prominent than the same-day effects. Given the null results, I

[78] First stage: the impact of abnormal shutdowns on followees' being offensive. Second stage: the impact of followees' being offensive on followers' being offensive. Reduced form: the impact of abnormal shutdowns on followers' being offensive. Followees of authors in Table S9.3 are 100% offensive, while tweet authors in Table 10 may not be offensive, so the first stage is stronger.

[79] Another reason for focusing on Facebook SCI and county-pair level aggregated results is the availability of crime data. I am not able to track individual users' locations given the incomplete location reporting, so crime effects are inconsistently observed in these followers' areas. Some users are not in the U.S., which further increases the difficulty of getting available crime records with similar data quality.

conclude that my earlier findings in Table 4 are not driven by offenders' observing hate speech online, but due to their direct exposure to local air pollution.[80]

Column (4) of Table 11 suggests the number of non-hate crimes increases by 0.3 (4.5%) as connectedness increases by 1 unit. Given my identifying assumption, the increase in crimes only results from the online network and the 1.3%-3.3% increase in aggressive tweets.[81] The result is consistent with existing evidence that reading offensive content invokes violence. In Panel B, the effects are more striking on the following day, with a 5.3% increase in crimes. While local responses are short-lasting, the spillover effects in other counties last for more than one day. My findings on local crime effects in Table 5 result from both air pollution exposure and online emotion contagion.

NIBRS data has national coverage and is comparable across time and places. Apart from national records, 14 cities' police department provides citywide 911 call records. While my analysis is less likely to have data quality concerns and inconsistent reporting over time, as shown in Table S9.22 Panel C and Figure S3.10, I use 911 call records as an alternative data source in the non-local crime analysis. For these 14 cities, there are 1,642,666 events from the NIBRS dataset and 2,458,069 events from 911 call records 2014-2019.[82] If I assume that city police services include NIBRS and have better data quality, I keep other counties' NIBRS reports, drop these 14 cities' NIBRS and add their 911 records. Results in Table S9.14 display similar estimates. Distant but online connected counties observe 0.34 (4.3%) more crimes on distant pollution shock days, compared with distant but non-connected counties. This practice confirms the robustness of my results and rules out the data quality concerns.[83]

How to compare non-local crime responses with local responses? On abnormal outage days, tweets and crimes in local and non-local areas respond to the pollution spike. In each local county-day, the number of offensive tweets increases by 33.56, 24.4% relative to the mean. The number of racist tweets increases by 5.62 (30.1%). The number of non-hate crimes increases by 5.18 (12.1%). The number of hate crimes and anti-Black hate crimes increases by 0.018 and 0.018, 22.5% and 59.1% respectively. In each non-local county-day where $Connected = 1$, compared with those with $Connected = 0$, the number of offensive tweets increases by 0.46 (3.3%).[84] The number of racist tweets increases by 0.063 (1.3%). The number of non-hate

[80] Black people are located close to refinery plants, shown in Table S8.7. Similarly, papers document racial segregation at the geographic scale in the U.S. (e.g. Reardon et al., 2008; Hall et al., 2019). My non-local findings on hate crimes against Black people show a large point estimate with low precision, probably due to low power and small Black populations in some counties. To address this concern, I drop counties in the bottom quartile of the Black population and only use counties in the top quartile. In Table S9.26, both practices generate larger point estimates and higher precisions of the interaction term, but both are still statistically insignificant or reach a borderline significance level of 10 percent.

[81] See Section 7.2 for explanations on the large estimated elasticity of tweets to crimes.

[82] 14 cities include: Los Angeles CA, Chicago IL, Phoenix AZ, San Antonio TX, Montgomery MD, San Jose CA, San Francisco CA, Seattle WA, Detroit MI, Baltimore MD, Tucson AZ, Mesa AZ, Sacramento CA, New Orleans LA. Crime data is available to download from each city's government website.

[83] Similar to Table S7.18, I use the union of counties with at least one hate crime or at least one non-hate crime reported 2014-2019 and fill non-reporting county-days with zero events. With the same sample sizes, Table S9.25 shows hate crimes against all race groups or Black groups have insignificant increases in response to remote pollution spikes. In contrast, non-hate crimes show positive and precise estimates on $UnplannedShut_{jt} \times Connected_{ij}$.

[84] 0.46 and 0.063 are calculated based on point estimates in Table 10. As non-local tweets are collected for five

crimes increases by 0.063 (1.3%). Despite statistically insignificant estimates, the number of hate crimes and anti-Black hate crimes increase by 0.0001 and 0.001, 1.5% and 33.3%. Taking the whole U.S. together, there are 93 with and 2236 counties without at least one refinery plant,[85] so the total crime burden in local and non-local areas in the U.S. is 484 and 154 if all refineries experience abnormal outages on that day. Nationally, the non-local effects are 31.8% as the local effects, and ignoring the geographical spillovers underestimates crime effects by 24.1%. If we further consider the transboundary effect, the non-local effects would be even higher as people have not only domestic friends but also online networks in other countries. Also, higher-order connections or the spillover of spillover would add magnitude and geographical range to the crime effect. Therefore, my estimates of 31.8% serve as the lower bound of non-local effects.

What could my empirical findings imply regarding emission tax and social interaction tax in Section 3? Based on my model, optimal emission tax minus d depends on connectedness, the partial derivative of n with respect to P, the total number of interactions, the marginal cost of receiving negative interactions, and the marginal benefit of receiving positive ones. While the last term is not covered, the former four are estimated in my empirical analysis. Assuming β_1 is zero, my results (24% of d) focusing on crime effects of negative interactions provide the lower bound of optimal emission tax. When it comes to optimal social tax, my model suggests it depends on the absolute value of n rather than its change with pollution. Although negative tweets serve as a signal for high n, my empirical findings do not cover all channels of physical and online interactions and could hardly generate implications for n. Future studies could estimate non-crime effects and benefits of social interactions, as well as regional baseline proportion of harmful contact, to provide other empirical estimates on optimal emission and social taxes.

9.4 Results using Twitter SCI

As discussed in Section 9.1, I focus on bad authors near ten refinery plants and scrape 969,961 followee-follower links so as to validate the Facebook SCI measure. The correlation test in Table S9.1 indicates that as #followers increase by 100, the connectedness index increases by 0.16 units. As a robustness check, I replace $Connected_{ij}$ in equation (2) with Twitter follower-based SCI and estimate the following equation:

$$Y_{jt} = \beta_1 UnplannedShut_{it} \times \#Followers_{ij} + Time_t + County_j + \varepsilon_{jt} \tag{5}$$

where $\#Followers_{ij}$ is the number of followers in county j that are following bad author in county i. $UnplannedShut_{it}$ is coded based on these ten plants and equals one if one of these ten plants i experiences an abnormal shutdown on the day t. Y_{jt} and other time and county fixed effects are the same as those in equation (2).

Since I scraped the top ten plants' follower information, I fill the other 91 plant i's $\#Followers_{ij}$

minutes on each day, I scale up the results by 288 (=60 minutes/hour ÷ 5 minutes collected × 24 hours/day).

[85] In the FBI crime dataset, there are 2329 counties with at least one crime event 2014-2019, so I ignore other 677 counties without any crime reported. Besides, 101 refinery plants in my sample are located in 93 different counties.

as zeros. Therefore, results in this section should be interpreted as the impact of the top ten plants' pollution shock on these ten plants' connected counties. This does not affect my identification on β_1. As other 91 plants' abnormal shutdown timing is exogenous, other event dummies are included in ε_{jt} and should not be correlated with current treatment term $UnplannedShut_{it} \times \#Followers_{ij}$.

Table 12 reports estimates on β_1 using offensive tweets, racist tweets, and non-hate crimes at the recipient-day level as dependent variables. As the number of followers that follow bad authors near refineries increase by 100, the number of offensive tweets posted in followers' counties increases by 0.45%; the number of racist tweets increases by 0.20%; the number of non-hate crime events increases by 0.88%. Given the correlation result in Table S9.1, magnitudes using Twitter follower-based SCI are comparable and larger than the main results in Section 9.2 and 9.3.[86] The elasticity of crime increase over tweet increase is also larger. A potential explanation is that the Facebook SCI data is constructed based on general online users. In contrast, my Twitter SCI is constructed based on bad authors' followers that tend to be a more aggressive subpopulation.

9.5 Possible threats to identification

I attribute the observed increase in faraway counties' aggressive activities to online connections. Though the sample only includes geographically distant areas that are 150km away from any refinery plant, spillover effects might be driven by the long-range cross-county pollution flow. Therefore, I conduct two additional checks to rule out the pollution dispersion channel. First, I use air pollution in geographically distant counties as the dependent variable and estimating equation (2). Results in Table S9.4 show air pollution does not respond to distant pollution shocks. Counties that are closely or loosely connected to polluters are exposed to a similar amount of pollution, which confirms that pollution itself could not drive the difference in aggressive behaviors.

The second check is controlling for the pollutant dispersion matrix. The EPA provides air quality modeling under the Cross-State Air Pollution Rule (CSAPR).[87] The model specifies ozone contribution values between state recipients and county polluters. I construct a similar connectivity matrix based on the contribution values:

$$Dispersion_{ij} = ln(\frac{Contribution_{ij}}{\sum_{k=1}^{N} Contribution_{kj}} + 1)$$

$$\begin{aligned} Y_{jt} = \beta_1 UnplannedShut_{it} \times Connected_{ij} + \beta_2 UnplannedShut_{it} \times Dispersion_{ij} \\ +\gamma OutageAnyPlant_t + Time_t + County_j + \varepsilon_{jt} \end{aligned} \tag{6}$$

[86] #Followers' increase by 100 is equivalent to connectedness index's increase by 0.16 unit. Results in Table 10 and 11 show as the connectedness index increases by 0.16 units, aggressive tweets increase by 0.21-0.53% and crimes increase by 0.72%.

[87] Detailed modeling and output data could be found at: https://www.epa.gov/csapr/final-cross-state-air-pollution-rule-update

where $Contribution_{ij}$ is the amount of ozone flow from emitter county i to receiver county j. $Dispersion_{ij}$ is the pollution connectedness index at the county pair level. This captures how much pollution is transported from county i relative to the total amount of pollution residents in county j is exposed to. Results reported in Table S9.5 show estimates on $UnplannedShut \times Connected$ are robust with additional controls. This suggests online connections still explain the increase in offensive and racist tweets after controlling for pollutant dispersion.

Another confounding channel is physical interaction. Local crimes may be spread to non-local areas through human movements and induced physical contact. This is not likely to happen in my case, given the distance threshold imposed and the same-day effect observed. I still conduct three additional checks to control for physical interactions. The first measure comes from the ACS mobility data. It asks respondents whether they lived in the same residence one year ago. Those living in a different residence would be collected their locations of previous and current residences.[88] In each year t, there are $Inflow_{ijt}$ individuals moving from county i to county j, and $Outflow_{ijt}$ individuals moving from j to i. Since potential physical spillovers from plants to recipients may threaten my analysis, I focus on $Inflow_{ijt}$ from refinery county i to recipient county j. The estimation equation is specified as:

$$Inflow_{ij} = ln[\frac{1}{6}\sum_{t=2014}^{2019}(Inflow_{ijt}) + 1]$$

$$\begin{aligned} Y_{jt} = \beta_1 UnplannedShut_{it} \times Connected_{ij} + \beta_2 UnplannedShut_{it} \times Inflow_{ij} \\ +\gamma OutageAnyPlant_t + Time_t + County_j + \varepsilon_{jt} \end{aligned} \tag{7}$$

The second measure of physical interaction is based on the Facebook colocation map which measures the probability that two individuals from county i and county j are found in the same location at the same time.[89] On each week t, there are $Population_{it}$ individuals from county i meeting $Population_{jt}$ individuals from county j. I take averages and collapse data from the county-pair-week level to the county-pair level over March 2020 to February 2022. I use the following method to control for physical connectedness:

$$Colocation_{ij} = ln(Population_i + Population_j + 1)$$

$$\begin{aligned} Y_{jt} = \beta_1 UnplannedShut_{it} \times Connected_{ij} + \beta_2 UnplannedShut_{it} \times Colocation_{ij} \\ +\gamma OutageAnyPlant_t + Time_t + County_j + \varepsilon_{jt} \end{aligned} \tag{8}$$

Coefficients from estimating equation (7) and (8) are displayed in Table S9.6 and S9.7.

[88] Data is available since 2005 from this link: https://www.census.gov/topics/population/migration/guidance/county-to-county-migration-flows.html

[89] Colocation maps are used by epidemiological papers to understand disease outbreaks and spread (e.g. Selinger et al., 2021; Shepherd et al., 2021; Fritz and Kauermann, 2022). The dataset is available to researchers upon request: https://dataforgood.facebook.com/dfg/tools/colocation-maps

Estimates on β_2 are positive and statistically significant, suggesting physical interactions indeed increase offensive behaviors on social media. Estimates on β_1 remain still strongly robust after controlling for physical connections, and R^2 is quite similar with and without the additional control.

Thirdly, I use geographic distance between counties to control for physical interactions. Geographic distance tends to be negatively correlated with the probability of physical contact. I replace $Colocation_{ij}$ in equation (7) with distance between county i and j. Results in Table S9.8 show very similar estimates on β_1, ruling out potential confounds that are positively correlated with geographic distances.

In addition, I control for potential information spread on traditional media. Existing studies reviewed in Section 2.1 document the impact of newspapers and other traditional media on physical crimes. The concern arises that if newspapers report local crimes near polluters, remote consumers of traditional media read crime incidents and commit crimes. To test this concern, I use media circulation data for all U.S. active newspapers and magazines from the Media Intelligence Center.[90] I construct a similar connectivity matrix to link subscribers' location and newspapers' headquarter.[91] Adding this additional connectivity matrix generates strongly robust results, as shown in Table S9.9.

In addition, one other potential confounder is oil price. Existing studies document a correlation between refinery outages and oil product prices in downstream markets (e.g. EIA, 2007; Chesnes, 2015), so non-local crimes may be driven by oil price changes rather than social media connections. Therefore, I obtain prices of gasoline products and related financial instruments and check price responses to outage schedules. Spot prices and future prices are from the EIA website at the national day level.[92] I use Conventional Gasoline spot price, RBOB Regular Gasoline spot price, and NYMEX futures price as dependent variables. I also obtain retail weekly gasoline prices at the district, state, or city level. Results from estimating equation (1) are reported in Table S9.15.[93] Oil prices have ambiguous responses to refinery outage schedules, probably due to the low spatial resolution or time frequency of oil price data. In general, refinery outages generate national shocks to downstream markets, and tend to increase prices of gasoline products and related financial instruments. Given the concern that pollution shock only affects distant counties through the channel of oil prices, I add oil prices as additional controls and re-estimate equation (2). As is shown in Table S9.16, estimates on $UnplannedShut \times Connected$ remain positive and significant, and have similar magnitudes as those in Table 10. After controlling for variations in oil prices, geographically distant but online connected areas still witness more

[90] The dataset includes 1429 active news media in total. Both print and digital circulation are covered. Details could be found on the Alliance for Audited Media website: https://auditedmedia.com/analysis-and-training/understanding-print-circulation

[91] Specifically, $Circulation_{ij}$ is calculated as the proportion of subscribers in county j out of all subscribers of newspaper n in county i. I use the simple average to aggregate multiple newspapers in each county.

[92] Spot price data is available here: https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm. Future price data link: https://www.eia.gov/dnav/pet/pet_pri_fut_s1_d.htm

[93] For spot and future prices, I stack the same outcome variables at the national-day level for all plants on that day. Similar for gasoline prices, I code the same value as outcome variables for all plants in that district on that week.

offensive content online due to refinery pollution shocks. Therefore, my findings on geographic spillovers are not entirely driven by downstream oil markets or gasoline-related financial prices.

Moreover, I construct a variable that captures the initial connectivity of recipient counties with refineries. In the spirit of Borusyak and Hull (2021), while the timing of unexpected shutdowns is random, whether counties are loosely or closely connected to refineries is pre-determined. Consequently, treatment variable $UnplannedShut_{it} \times Connected_{ij}$ involves non-random exposure to exogenous shocks. Borusyak and Hull (2021) suggests a solution based on simulating counterfactual shocks that could have been realized. I reshuffle the date of abnormal shutdowns within or across plants in my sample.[94] Table S9.17 shows that the coefficients are robust to applying the recentered treatment. I conclude that my findings on distant aggressive activities are well identified.

9.6 Inward spillovers

In Section 7.1, local tweet responses to refinery pollution shock are denoted by $y(z)$, where y are tweet outcomes and z is an indicator for local abnormal outages. To identify the causal impact of z on y, I need to assume the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1978) or individualistic treatment response (ITR) (Manski, 2013) that restricts an individual's outcome may vary only with her own treatment, not those of other members of the population. As social interactions are common within and across areas, outcomes vary with the treatment of others, which is confirmed by results in Section 9.2. Based on (Haan, 2017), an extended definition of treatment effect is the difference between two mean potential outcomes: the mean outcome we would observe if abnormal-induced pollution shocks occured in all refineries, and the mean outcome we would observe if no refineries receive pollution shocks:

$$E[y_i(z_i = 1, z_{-i} = 1)] - E[y_i(z_i = 0, z_{-i} = 0)]$$

The treatment effect could be identified under the monotone spillover assumption (MSP):

$$E[y_i(z_i = 0, z_{-i} = 0)] \leq E[y_i(z_i = 0, z_{-i} = z^{J-1})] \leq E[y_i(z_i = 0, z_{-i} = 1)]$$

$$E[y_i(z_i = 1, z_{-i} = 1)] \geq E[y_i(z_i = 1, z_{-i} = z^{J-1})] \geq E[y_i(z_i = 1, z_{-i} = 0)]$$

where $z_{-i} = z^{J-1}$ specifies the pollution shock actually experienced by other plants. This assumption accounts for positive spillovers but not for negative spillovers. In other words, aggressive activities are amplified through social networks but not mitigated. Under MSP, my main estimate serves as the lower bound of the extended treatment effect, i.e. $E[y_i(z_i = 1, z_{-i} = z^{J-1})] - E[y_i(z_i = 0, z_{-i} = z^{J-1})] \leq E[y_i(z_i = 1, z_{-i} = 1)] - E[y_i(z_i = 0, z_{-i} = 0)]$.

[94] In other words, I do two different randomizations. First, I keep the actual number of abnormal shutdowns of each plant and randomize the schedule within each plant over the six years, i.e. at the time level. Second, I take all plants together, keep the total number of shutdowns, and randomize the schedule at both plant and time levels.

I also use an alternative estimation strategy proposed by (James and Smith, 2020) to account for the inward spillover. Based on equation (1), I further add $UnplannedShut\_Distant_{jt} \times Connected_{ij}$ to control other connected plants' treatment. Here $UnplannedShut\_Distant_{jt}$ is other plant j's outages, and $Connected_{ij}$ is the connectedness between outcome plant i and connected plant j. I code this spillover treatment measure similar to that in equation (2), using either the absolute or relative measure to define $Connected_{ij}$, either the maximum or the average value to aggregate multiple connected plants j. This practice estimates the impact of own plant's pollution shock, conditional on connected plants' shocks. In Table S9.18, I find a positive, significant, and slightly smaller increase in surrounding offensive tweets. The change in point estimates on $UnplannedShut$ is not statistically significant compared with that in Table 3.[95] Specifically, the number of offensive tweets increases by 33.1 and 33.6 with and without connectedness controls, 24.0% and 24.4% relative to the main. The number of racist tweets increases by 5.4 and 5.6, 29.0% to 30.1% of the average.

To account for inward spillover concerns in crime results, I also add $UnplannedShut_{jt} \times Connected_{ij}$ in local crime analysis. Results in Table S9.20 are not statistically different with those in Table 5. Counties where refineries are located experience 4.9 (11.5%) more non-hate crimes on the abnormal outage days. The equality test of two coefficients on $UnplannedShut$ has a z-score equal to 0.0812, and the corresponding p-value is 0.4676.

Furthermore, I use a subset of counties whose connectedness indexes are in the lowest quartile. This practice is to reduce the concern of inward spillovers. Results on tweets and crimes are summarized in Table S9.19 and S9.21. Estimates on $UnplannedShut$ are strongly robust, which confirms reverse spillovers do not affect the local effect of pollution on local aggressive behaviors. The magnitudes are slightly larger than the main estimates. Online isolated areas may have limited across-region connections and are more affected by within-region events and internal polluting activities.

10 Possible mechanisms for local aggressive behaviors

My empirical findings suggest that high air pollution has significantly adverse effects on aggressive behaviors, including more offensive posts on Twitter and more crimes. In this section, I shed light on the mechanisms of my local findings. I examine two primary hypotheses regarding the underlying mechanisms: (i) psychological effects - bad air quality lowers impulse control and increases the underlying anger and aggression; (ii) policing activities - police governance is negatively affected by pollution, which in turn increases people's offensive behaviors.

[95] I conduct a z-test to compare estimate on $UnplannedShut$ in Table S9.18 and Table 3. Using offensive tweets as dependent variables, z-score is 0.0326, and corresponding p-value is 0.4869. For racist tweets, z-score and p-value are 0.0554 and 0.4779 respectively. Both tests could not reject the null hypothesis that two estimated coefficients are equal.

10.1 Psychological explanations

In psychology, impulse behavior is when people act quickly without thoughts to the consequences. Five behavioral stages characterize impulsivity: impulse, growing tension, pleasure on acting, relief from the urge, and potential guilt (Wright et al., 2012). Fortunately, people are able to restrain most impulsivity by impulse control - the ability to control and regulate impulses, emotions, desires, and performances (Tice et al., 2001). If a person has trouble controlling emotions or behaviors, impulse control issues or disorders occur. The latter category includes including pathological gambling, compulsive shopping, hypersexuality, compulsive eating, compulsive medication use, hoarding, kleptomania, and impulsive smoking (Voon et al., 2011). Expressed or observed behavior is a function of the underlying feeling and the level of impulse control. I hypothesize that air pollution decreases impulse control, so people post more offensive tweets and commit more crimes, even if their underlying level of anger and aggression is not affected. The alternative hypothesis is that air pollution increases anger and aggression. Even with the same impulse control, people still post more offensive tweets and commit more crimes.

I use expenditure on snacks (also used by Haws et al. (2015)) and visits to casinos (Ledgerwood and Petry, 2015) to indicate impulse control levels. They come from the Nielsen consumer panel and the Safegraph foot traffic data described in Section 4. These measures respond to impulse control but are not affected by underlying anger. Results in Table S10.1 show consumption on some snacks increase, while visits to gambling places do not respond to outage schedule despite positive estimates. Specifically, expenditure on desserts and ice cream increase by \$0.3 and \$0.6 in the county where abnormal outages occur, equivalent to 16.8% and 7.2% of the average expenditure. Therefore, lower impulse control could explain observed more offensive tweets and physical violence to some extent.[96]

10.2 Police governance

The second potential mechanism lies in police activities. Police presence reduces crime (Di Tella and Schargrodsky, 2004; Draca et al., 2011), and people may be more offensive after observing fewer police on patrol due to air pollution. The supply side response may partly explain the observed higher aggression. Existing papers find police deployment does not respond to air pollution (Bondy et al., 2020). On other environmental conditions, Obradovich et al. (2018) and Heilmann et al. (2021) have shown that temperature leads to fewer food safety inspections and vehicle stops, and more Code 6 investigations and arrests. Despite an effort shift, the overall police activities and deterrence do not respond to adverse weather conditions.

To test police efforts' response to outage schedules, I obtain data from the Stanford Open

[96] I am not able to test the underlying level of anger and aggression given lower impulse control. In other words, any observed quarrels or conflicts that respond to the underlying level of rage would increase with lower impulse control. One may infer a higher underlying anger by comparing the magnitude of effects in Table S10.1 with that in Table 3 and 4. However, the functional form of expressed offensive behaviors is not known, so a decrease in impulse control may mechanically lead to a small increase in snack expenditure and a large increase in criminal activities. Therefore, I conclude that lower impulse control could explain observed air pollution effects to some degree and do not attempt to test the underlying anger.

Policing Project developed by (Pierson et al., 2020). It covers 21 state patrol agencies and 35 municipal police departments. For all individual-level stop records, I am able to observe date and time, while coordinates and officer identifiers are incompletely documented. I use the number of tickets and the number of officers at the county-day level to measure police activities.

Reduced form results in Table S10.2 show neither the number of tickets nor the number of officers responses to outages and their induced air pollution change. Despite low precision, the estimate on $UnplannedShut$ in Column (2) is large and negative, indicating insignificantly lower efforts in traffic governance when pollution and other crimes increase. In contrast, the estimate in Column (1) is positive and imprecise. My findings conclude that low police governance is not a mechanism of the observed more offensive activities and are consistent with existing papers documenting little response of police deployment to adverse environmental factors.

11 Conclusion

The rise of social media and its increasingly broad usage have highlighted the importance of understanding users' online activities, their driving factors, and consequences. Typically, social media is shaped by internal characteristics and external socioeconomic circumstances. This paper contributes to the latter aspect from the environmental direction. When ambient air pollution goes up, surrounding residents complain more about pollution and write more offensive and racist posts on Twitter. Changes in social media activities not only serve as direct impacts of pollution, but also generate ripple effects on the mental health of followers and friends, as well as crimes and violence.

Furthermore, this paper contributes to environmental economics by providing a new natural experiment for air pollution by exploiting polluters' outage schedules. The identification is not restricted by weather conditions, geographic or time ranges, and frequently occurs in the refining sector worldwide. Using continuous pollution readings from satellite products and rich information on crime and health outcomes, my calculation reveals that a one standard deviation increase in surrounding air pollution leads to 0.16 more hate crimes against black people and 0.23 more hospital visits per thousand population on each day. My findings are comparable and slightly larger than existing studies (e.g. Schlenker and Walker, 2015; Deryugina et al., 2019) that mainly focus on other criteria air pollutants based on ground-based pollution measurements. The high impacts of VOCs make this pollutant meaningful criteria for future research, detailed readings from ambient monitors, and environmental policies.

Adverse effects of refineries' abnormal operations call for advanced technologies from operators and awareness from the public. Unexpected shutdowns take place every year in refineries worldwide, so advancement is much needed to control abnormal operations and minimize their impacts. My findings are applicable not only in the U.S., but also in oil refineries elsewhere. Within the U.S., existing ambient monitors are far from polluters and insufficient to capture their surrounding air quality. This results in little public knowledge of polluting events and surrounding residents' similar outdoor activities on days with harmful air quality. Given the

fact that air quality monitoring in developed countries is more systematically established and maintained than that in developing countries (Johnson et al., 2011; Gulia et al., 2020), this paper provides vital insights on the need for better measurements and pollution reporting.

Section S11 proposes a potential policy solution - lowering utilization rates - that could reduce the frequency and the severity of abnormal shutdowns. The costs incurred by refinery plants due to the lost oil products and revenue are lower than the considerable morbidity-related benefits. Therefore, current conditions with no enforceable regulation are far from efficient, and social welfare has much room for improvement. This paper opens avenues for future research to explore more cost-effective environmental policies. It would be helpful to explore other policy options to internalize the external cost of oil refining, including command-and-control approaches like my proposed policy and market-based instruments.

On a broader level beyond air pollution, this paper provides evidence on multiple externalities that exacerbate the impact of environmental degradation. The externality on directly exposed units could further generate adverse consequences on other exposed and unexposed ones. The presence of externalities' interactions reinforces the importance of regulating and minimizing hazards, monitoring and disclosing information, improving awareness and avoidance in environmental discussions.

My findings are generalizable to other environmental stressors that could induce aggressive behaviors. One important direction of future research is climate change. My current setting focuses on short-term effects in response to one-day abnormal operations, while climate change is a long-term problem. Future works could explore climate projection and predicted crime and social media responses. More weather events due to climate change could induce more abnormal operations or other emission events. Given my results, both local and non-local effects will be turned on due to more pollution. At the same time, weather and climate variables themselves could be environmental drivers that induce aggressive behaviors online and offline, so the finding and framework in this paper could be generalizable to study weather-induced toxic tweets and in turn crimes. One difference that may be interesting is the magnitude of findings. With a change in steady state and baseline weather conditions, the effect of toxic tweets or violent behaviors may be different if nonlinearity exists.

Environmental implications aside, I show policymakers should regulate social activities in regions with many negative interactions, either online or physically. Using a market-based tool based on my model, social tax increases with connectedness within and across regions, the proportion of negative interactions in each region's baseline contact pool, the marginal cost or mental damage of receiving harmful interactions, and the marginal benefit of receiving positive interactions. While the last term is not covered, the first three are empirically estimated in this study, despite only focusing on crime-related costs of negative interactions. Therefore, my findings serve as a lower bound of the optimal social interaction tax.

Focusing on online interactions, my findings quantify the adverse effect of aggressive social media content on crimes and underscore the power of online interactions. Despite the positive

emotional support and social activism, negative interactions on social media, like harassment, cyberbullying, violence promotion, and discrimination have not been emphasized enough given the speed at which social media has been adopted into our lives. This paper highlights new challenges related to regulating online activities, designing appropriate policies, and counteracting its damaging effect. With connectivity measures across countries, future research could test whether my identified relationship between social media and physical crimes in the U.S. could be generalized to other countries.

I focus on the U.S. to empirically test the impact of online toxic content as the government is prevented from regulating or restricting speech. Future works could use other countries with some limitations and censorship on social media. By comparing the across-country difference, studies could explore the role of intermediation in the spread of social media on real-world violence.

One possible policy implication lies in content diffusion or popularity penalization of aggressive content. In other words, free speech is not the same as free reach. Currently, online search engines and social media platforms are able to order and rank results displayed on their pages and feeds. This underlying, invisible, and powerful algorithm determines what content billions of internet users read, watch and share next. The algorithms are able to predict what we want to see, but they don't differentiate what is propaganda and what is not, what is fake news and what is fact-checked (Diresta, 2018). To make matters worse, incendiary, controversial, and polarizing posts receive more reactions, and tend to be amplified by the engagement-based diffusion algorithm. With a few taps and reactionary buttons, harmful content could be amplified by online platforms and algorithms into harmful cascades. Therefore, we need transparency and improvement in the algorithmic curation, incentives, and outcomes.

References

Aghababaei, S. and M. Makrehchi (2016). Mining social media content for crime prediction. In *2016 ieee/wic/acm international conference on web intelligence (wi)*, pp. 526–531. IEEE.

ALA (2022). State of the air 2022. Technical report, American Lung Association.

Allen, J. L., K. Conrad, G. Oberdörster, C. J. Johnston, B. Sleezer, and D. A. Cory-Slechta (2013). Developmental exposure to concentrated ambient particles and preference for immediate reward in mice. *Environmental Health Perspectives 121*(1), 32–38.

Almond, D. and X. Du (2020). Later bedtimes predict president trump's performance. *Economics Letters 197*, 109590.

Almond, D., X. Du, and A. Vogel (2022, 03). Reduced trolling on russian holidays and daily us presidential election odds. *PLOS ONE 17*(3), 1–19.

Alvarez, R. A., D. Zavala-Araiza, D. R. Lyon, D. T. Allen, Z. R. Barkley, A. R. Brandt, K. J. Davis, S. C. Herndon, D. J. Jacob, A. Karion, E. A. Kort, B. K. Lamb, T. Lauvaux, J. D. Maasakkers, A. J. Marchese, M. Omara, S. W. Pacala, J. Peischl, A. L. Robinson, P. B. Shepson, C. Sweeney, A. Townsend-Small, S. C. Wofsy, and S. P. Hamburg (2018). Assessment of methane emissions from the u.s. oil and gas supply chain. *Science 361*(6398), 186–188.

Ambrey, C. L., C. M. Fleming, and A. Y.-C. Chan (2014). Estimating the cost of air pollution in south east queensland: An application of the life satisfaction non-market valuation approach. *Ecological Economics 97*, 172–181.

Amitai, Y., Z. Zlotogorski, V. Golan-Katzav, A. Wexler, and D. Gross (1998). Neuropsychological impairment from acute low-level exposure to carbon monoxide. *Archives of neurology 55*(6), 845–848.

Anderson, C. A. and B. J. Bushman (2002). Human aggression. *Annual Review of Psychology 53*(1), 27–51. PMID: 11752478.

Anderson, M. L. (2019, 10). As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality. *Journal of the European Economic Association 18*(4), 1886–1927.

Andison, F. S. (1977). Tv violence and viewer aggression: A cumulation of study results 1956–1976. *Public Opinion Quarterly 41*, 314–331.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*(1), 727–753.

Arceo, E., R. Hanna, and P. Oliva (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal 126*(591), 257–280.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy 7*(2), 181–198.

Austin, S. B., S. H. Liu, and N. Tefft (2018). Could a tax on unhealthy products sold for weight loss reduce consumer use? a novel estimation of potential taxation effects. *Preventive Medicine 114*, 39–46.

Austin, W., S. Carattini, J. G. Mahecha, and M. Pesko (2020). Covid-19 mortality and contemporaneous air pollution.

Azzimonti, M. and M. Fernandes (2022). Social media networks, fake news, and polarization. *European Journal of Political Economy*, 102256.

Bagilet, V. and L. Zabrocki (2022, March). Unconfounded but inflated causal estimates. *Working paper*.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy 126*(6), 2224–2276.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018, August). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives 32*(3), 259–80.

Bailey, M., D. M. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2019, May). Peer effects in product adoption. Working Paper 25843, National Bureau of Economic Research.

Bailey, M., T. Kuchler, D. Russel, B. State, and J. Stroebel (2020). The determinants and effects of social connectedness in europe. CESifo Working Paper No. 8310, Available at SSRN: https://ssrn.com/abstract=3759016.

Bandura, A. (1973). Aggression: A social learning analysis. *Prentice-Hall*.

Bandura, A. (1983). Psychological mechanisms of aggression. *Aggression: Theoretical and Empirical Reviews*.

Barnett, A. H. (1980). The pigouvian tax rule under monopoly. *American Economic Review 70*(5), 1037–41.

Baron, R. A. and P. A. Bell (1976). Aggression and heat: the influence of ambient temperature, negative affect, and a cooling drink on physical aggression. *Journal of personality and social psychology 33*(3), 245.

Barrows, G., T. Garg, and A. Jha (2019, December). The health costs of coal-fired power plants in india. *IZA Discussion Paper Series* (12838).

Barwick, P. J., S. Li, D. Rao, and N. B. Zahur (2018, June). The healthcare cost of air pollution: Evidence from the world's largest payment network. Working Paper 24688, National Bureau of Economic Research.

Baurick, T., L. Younes, and J. Meiners (2019). Welcome to "cancer alley" where toxic air is about to get worse. *ProPublica*.

Baylis, P. (2020). Temperature and temperament: Evidence from twitter. *Journal of Public Economics 184*, 104161.

Bento, A., D. Kaffine, K. Roth, and M. Zaragoza-Watkins (2014, August). The effects of regulation in the presence of multiple unpriced externalities: Evidence from the transportation sector. *American Economic Journal: Economic Policy 6*(3), 1–29.

Berkowitz, L. (1984). Some effects of thoughts on anti- and prosocial influences of media events: A cognitive-neoassociation analysis. *Psychological Bulletin 95*(3), 410–427.

Berman, E. and L. Bui (2001, 02). Environmental regulation and productivity: Evidence from oil refineries. *The Review of Economics and Statistics 83*, 498–510.

Beurel, E. and R. Jope (2014). Inflammation and lithium: clues to mechanisms contributing to suicide-linked traits. *Translational psychiatry 4*(12), e488–e488.

Bhuller, M., T. Havnes, E. Leuven, and M. Mogstad (2013, 04). Broadband Internet: An Information Superhighway to Sex Crime? *The Review of Economic Studies 80*(4), 1237–1266.

Birger, M., M. Swartz, D. Cohen, Y. Alesh, C. Grishpan, and M. Kotelr (2003). Aggression: the testosterone-serotonin link. *IMAJ-RAMAT GAN- 5*(9), 653–658.

Block, M. L. and L. Calderón-Garcidueñas (2009). Air pollution: mechanisms of neuroinflammation and cns disease. *Trends in neurosciences 32*(9), 506–516.

Boer, J. T., M. Pastor, J. L. Sadd, and L. D. Snyder (1997). Is there environmental racism? the demographics of hazardous waste in los angeles county. *Social Science Quarterly 78*(4), 793–810.

Bondy, M., S. Roth, and L. Sager (2020). Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists 7*(3), 555–585.

Borusyak, K. and P. Hull (2021, January). Non-random exposure to exogenous shocks: Theory and applications. Working Paper 27845, National Bureau of Economic Research.

Bovenberg, A. L. and L. H. Goulder (1996, September). Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses. *American Economic Review 86*(4), 985–1000.

Buchanan, J. (1969). External diseconomies, corrective taxes, and market structure. *American Economic Review 59*(1), 174–77.

Burkhardt, J. (2019). The impact of the Renewable Fuel Standard on US oil refineries. *Energy Policy 130*(C), 429–437.

Burkhardt, J., J. Bayham, A. Wilson, E. Carter, J. D. Berman, K. O'Dell, B. Ford, E. V. Fischer, and J. R. Pierce (2019). The effect of pollution on crime: Evidence from data on particulate matter and ozone. *Journal of Environmental Economics and Management 98*, 102267.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019, December). Social media and xenophobia: Evidence from russia. Working Paper 26567, National Bureau of Economic Research.

Card, D. and G. Dahl (2011, 04). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The quarterly journal of economics 126*, 103–43.

Carpenter, A. and M. Wagner (2019a). Environmental justice in the oil refinery industry: A panel analysis across united states counties. *Ecological Economics 159*, 101–109.

Carpenter, A. and M. Wagner (2019b). Environmental justice in the oil refinery industry: A panel analysis across united states counties. *Ecological Economics 159*, 101–109.

Chang, S., E. Pierson, P. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec (2021, 01). Mobility network models of covid-19 explain inequities and inform reopening. *Nature 589*, 1–6.

Chay, K. Y. and M. Greenstone (2003). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics 118*(3), 1121–1167.

Chen, C.-Y., A. C. Bonham, C. G. Plopper, and J. P. Joad (2003). Selected contribution: Neuroplasticity in nucleus tractus solitarius neurons after episodic ozone exposure in infant primates. *Journal of Applied Physiology 94*(2), 819–827.

Chen, S. and T. Li (2020). The effect of air pollution on criminal activities: Evidence from the nox budget trading program. *Regional Science and Urban Economics 83*, 103528.

Chen, S., P. Oliva, and P. Zhang (2018, June). Air pollution and mental health: Evidence from china. Working Paper 24686, National Bureau of Economic Research.

Chesnes, M. (2015). The impact of outages on prices and investment in the u.s. oil refining industry. *Energy Economics 50*, 324–336.

Coccaro, E. F., C. S. Sripada, R. N. Yanowitch, and K. L. Phan (2011). Corticolimbic function in impulsive aggressive behavior. *Biological psychiatry 69*(12), 1153–1159.

Colmer, J., I. Hardman, J. Shimshack, and J. Voorheis (2020). Disparities in pm_{2.5} air pollution in the united states. *Science 369*(6503), 575–578.

Cook, J., N. Newberger, and S. Smalling (2020). The spread of social distancing. *Economics Letters 196*, 109511.

Cotti, C., E. Nesson, and N. Tefft (2019). Impacts of the aca medicaid expansion on health behaviors: Evidence from household panel data. *Health Economics 28*(2), 219–244.

Crockett, M. J., A. Apergis-Schoute, B. Herrmann, M. D. Lieberman, U. Müller, T. W. Robbins, and L. Clark (2013). Serotonin modulates striatal responses to fairness and retaliation in humans. *Journal of neuroscience 33*(8), 3505–3513.

Dabbs Jr, J. M., T. S. Carr, R. L. Frady, and J. K. Riad (1995). Testosterone, crime, and misbehavior among 692 male prison inmates. *Personality and individual Differences 18*(5), 627–633.

Dahl, G. and S. DellaVigna (2009). Does movie violence increase violent crime? *The Quarterly Journal of Economics 124*(2), 677–734.

Davis, L. W. and E. Muehlegger (2010). Do americans consume too little natural gas? an empirical test of marginal cost pricing. *The RAND Journal of Economics 41*(4), 791–810.

de Chaisemartin, C. and X. D'Haultfoeuille (2020, September). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

Dell, M., B. F. Jones, and B. A. Olken (2012, July). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics 4*(3), 66–95.

DellaVigna, S. and M. Gentzkow (2010). Persuasion: Empirical evidence. *Annual Review of Economics 2*(1), 643–669.

DellaVigna, S. and E. La Ferrara (2015). Chapter 19 - economic and social impacts of the media. In S. P. Anderson, J. Waldfogel, and D. Strömberg (Eds.), *Handbook of Media Economics*, Volume 1 of *Handbook of Media Economics*, pp. 723–768. North-Holland.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019, December). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review 109*(12), 4178–4219.

Deschenes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us. *American Economic Journal: Applied Economics 3*(4), 152–185.

Deschenes, O., M. Greenstone, and J. S. Shapiro (2017, October). Defensive investments and the demand for air quality: Evidence from the nox budget program. *American Economic Review 107*(10), 2958–89.

Di Tella, R. and R. MacCulloch (2008). Gross national happiness as an answer to the easterlin paradox? *Journal of Development Economics 86*(1), 22–42.

Di Tella, R. and E. Schargrodsky (2004, March). Do police reduce crime? estimates using the allocation of police forces after a terrorist attack. *American Economic Review 94*(1), 115–133.

Diemer, A. and T. Regan (2020, November). No inventor is an island: social connectedness and the geography of knowledge flows in the US. CEP Discussion Papers dp1731, Centre for Economic Performance, LSE.

Diresta, R. (2018). Free speech is not the same as free reach.

Doremus, J. M., S. S. Stith, and J. M. Vigil (2019). Using recreational cannabis to treat insomnia: Evidence from over-the-counter sleep aid sales in colorado. *Complementary Therapies in Medicine 47*, 102207.

Draca, M., S. Machin, and R. Witt (2011, August). Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *American Economic Review 101*(5), 2157–81.

Drury, B., S. M. Drury, M. A. Rahman, and I. Ullah (2022). A social network of crime: A review of the use of social networks for crime and the detection of crime. *Online Social Networks and Media 30*, 100211.

EIA (2007, 03). Refinery outages: Description and potential impact on petroleum product prices.

Enikolopov, R., A. Makarin, and M. Petrova (2016, May). Social Media and Protest Participation: Evidence from Russia. CEPR Discussion Papers 11254, C.E.P.R. Discussion Papers.

Fan, M., G. He, and M. Zhou (2020). The winter choke: Coal-fired heating, air pollution, and mortality in china. *Journal of Health Economics 71*, 102316.

Farboodi, M., G. Jarosch, and R. Shimer (2020, April). Internal and external effects of social distancing in a pandemic. Working Paper 27059, National Bureau of Economic Research.

Faustman, W. O., D. L. Ringo, and K. F. Faull (1993). An association between low levels of 5-hiaa and hva in cerebrospinal fluid and early mortality in a diagnostically mixed psychiatric sample. *The British Journal of Psychiatry 163*(4), 519–521.

Felson, R. B. (1996). Mass media effects on violent behavior. *Annual Review of Sociology 22*(1), 103–128.

Flick, C. (2016). Informed consent and the facebook emotional manipulation study. *Research Ethics 12*(1), 14–28.

Fowler, J. H. and N. A. Christakis (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *BMJ 337*.

Fowler, S. L., A. Stylianou, R. Mousavi, S. Reid, and D. Zhang (2020). Detecting violent crime with gang social media postings.

Frankle, W. G., I. Lombardo, A. S. New, M. Goodman, P. S. Talbot, Y. Huang, D.-R. Hwang, M. Slifstein, S. Curry, A. Abi-Dargham, et al. (2005). Brain serotonin transporter distribution in subjects with impulsive aggressivity: a positron emission study with [11c] mcn 5652. *American Journal of Psychiatry 162*(5), 915–923.

Fried, A., J. Walega, P. Weibring, D. Richter, I. J. Simpson, D. R. Blake, N. J. Blake, S. Meinardi, B. Barletta, S. C. Hughes, J. H. Crawford, G. Diskin, J. Barrick, J. Hair, M. Fenn, A. Wisthaler, T. Mikoviny, J.-H. Woo, M. Park, J. Kim, K.-E. Min, S. Jeong, P. O. Wennberg, M. J. Kim, J. D. Crounse, A. P. Teng, R. Bennett, M. Yang-Martin, M. A. Shook, G. Huey, D. Tanner, C. Knote, J. Kim, R. Park, and W. Brune (2020, 12). Airborne formaldehyde and volatile organic compound measurements over the Daesan petrochemical complex on Korea's

northwest coast during the Korea-United States Air Quality study: Estimation of emission fluxes and effects on air quality. *Elementa: Science of the Anthropocene 8*(1). 121.

Fritz, C. and G. Kauermann (2020, 08). On the interplay of regional mobility, social connectedness, and the spread of covid-19 in germany.

Fritz, C. and G. Kauermann (2022). On the interplay of regional mobility, social connectedness and the spread of covid-19 in germany. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 185*(1), 400–424.

Gao, Y., Y. Chen, L. Mu, S. Gong, P. Zhang, and Y. Liu (2022, 04). Measuring urban sentiments from social media data: a dual-polarity metric approach. *Journal of Geographical Systems 24*.

Gavazza, A., M. Nardotto, and T. Valletti (2018, 06). Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies. *The Review of Economic Studies 86*(5), 2092–2135.

Gelman, A. and J. Carlin (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science 9*(6), 641–651. PMID: 26186114.

Gerber, A., J. Gimpel, D. Green, and D. Shaw (2011, 02). How large and long-lasting are the persuasive effects of televised campaign ads? results from a randomized field experiment. *American Political Science Review 105*, 135–150.

Gerber, A. S., D. Karlan, and D. Bergan (2009, April). Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics 1*(2), 35–52.

Giertz, S. (2006, 01). The elasticity of taxable income during the 1990s: A sensitivity analysis. *University Library of Munich, Germany, MPRA Paper*.

González-Guevara, E., J. C. Martínez-Lazcano, V. Custodio, M. Hernández-Cerón, C. Rubio, and C. Paz (2014). Exposure to ozone induces a systemic inflammatory response: possible source of the neurological alterations induced by this gas. *Inhalation toxicology 26*(8), 485–491.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277. Themed Issue: Treatment Effect 1.

Goulder, L. H. (1998). Environmental policy making in a second-best setting. *Journal of Applied Economics 1*(2), 279–328.

Graham, J. D., N. D. Beaulieu, D. Sussman, M. Sadowitz, and Y.-C. Li (1999). Who lives near coke plants and oil refineries? an exploration of the environmental inequity hypothesis. *Risk Analysis 19*(2), 171–186.

Greenstone, M. and R. Hanna (2014, October). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review 104*(10), 3038–72.

Greenstone, M., G. He, R. Jia, and T. Liu (2020, July). Can technology solve the principal-agent problem? evidence from chinaÂ's war on air pollution. Working Paper 27502, National Bureau of Economic Research.

Gulia, S., I. Khanna, K. Shukla, and M. Khare (2020). Ambient air pollutant monitoring and analysis protocol for low and middle income countries: An element of comprehensive urban air quality management framework. *Atmospheric Environment 222*, 117120.

Guo, J., L. Yang, R. Bie, J. Yu, Y. Gao, Y. Shen, and A. Kos (2019). An xgboost-based physical fitness evaluation model using advanced feature selection and bayesian hyper-parameter optimization for wearable running monitoring. *Computer Networks 151*, 166–180.

Haan, M. D. (2017, February). The Effect of Additional Funds for Low-ability Pupils: A Non-parametric Bounds Analysis. *Economic Journal 127*(599), 177–198.

Hall, M., J. Iceland, and Y. Yi (2019). Racial separation at home and work: Segregation in residential and workplace settings. *Population Research and Policy Review 38*(5), 671–694.

Hallinan, B., J. R. Brubaker, and C. Fiesler (2020). Unexpected expectations: Public reaction to the facebook emotional contagion study. *New Media & Society 22*(6), 1076–1094.

Haws, K., S. Davis, and U. Dholakia (2015, 01). Salad = success and fries = failure? conceptualizing and assessing self-control outcome measures in food decision-making research. *Journal of Consumer Behaviour 15*, n/a–n/a.

He, G., M. Fan, and M. Zhou (2016). The effect of air pollution on mortality in china: Evidence from the 2008 beijing olympic games. *Journal of Environmental Economics and Management 79*, 18–39.

He, J., H. Liu, and A. Salvo (2019, January). Severe air pollution and labor productivity: Evidence from industrial towns in china. *American Economic Journal: Applied Economics 11*(1), 173–201.

Heal, G. (2021, September). Empathy and the efficient provision of public goods. Working Paper 29255, National Bureau of Economic Research.

Heilmann, K., M. E. Kahn, and C. K. Tang (2021). The urban crime and heat gradient in high and low poverty areas. *Journal of Public Economics 197*, 104408.

Heim, B. (2009, 03). Structural estimation of family labor supply with taxes: Estimating a continuous hours model using a direct utility specification. *Journal of Human Resources 44*.

Herrnstadt, E., A. Heyes, E. Muehlegger, and S. Saberian (2021). Air pollution and criminal activity: Microgeographic evidence from chicago. *American Economic Journal: Applied Economics*.

Heyes, A. and M. Zhu (2019). Air pollution as a cause of sleeplessness: Social media evidence from a panel of chinese cities. *Journal of Environmental Economics and Management 98*, 102247.

Holtz, D., M. Zhao, S. G. Benzell, C. Y. Cao, M. A. Rahimian, J. Yang, J. Allen, A. Collis, A. Moehring, T. Sowrirajan, D. Ghosh, Y. Zhang, P. S. Dhillon, C. Nicolaides, D. Eckles, and S. Aral (2020a). Interdependence and the cost of uncoordinated responses to covid-19. *Proceedings of the National Academy of Sciences 117*(33), 19837–19843.

Holtz, D., M. Zhao, S. G. Benzell, C. Y. Cao, M. A. Rahimian, J. Yang, J. Allen, A. Collis, A. Moehring, T. Sowrirajan, D. Ghosh, Y. Zhang, P. S. Dhillon, C. Nicolaides, D. Eckles, and S. Aral (2020b). Interdependence and the cost of uncoordinated responses to covid-19. *Proceedings of the National Academy of Sciences 117*(33), 19837–19843.

Houweling, S., F. Dentener, and J. Lelieveld (1998). The impact of nonmethane hydrocarbon compounds on tropospheric photochemistry. *Journal of Geophysical Research: Atmospheres 103*(D9), 10673–10696.

Howard, J. W. (2019). Free speech and hate speech. *Annual Review of Political Science 22*(1), 93–109.

Hsiang, S., P. Oliva, and R. Walker (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy 13*(1), 83–103.

Hu, Z. (2020, November). Salience and households' flood insurance decisions. Mimeo, Available at SSRN: https://ssrn.com/abstract=3759016.

Huber, G. A. and K. Arceneaux (2007). Identifying the persuasive effects of presidential advertising. *American Journal of Political Science 51*(4), 957–977.

Hutto, C. and E. Gilbert (2015, 01). Vader: A parsimonious rule-based model for sentiment analysis of social media text.

Iganski, P. and S. Lagou (2015). Hate crimes hurt some more than others: Implications for the just sentencing of offenders. *Journal of Interpersonal Violence 30*(10), 1696–1718. PMID: 25201920.

Imelda (2018, May). Indoor air pollution and infant mortality: A new approach. *AEA Papers and Proceedings 108*, 416–21.

Ince, J., F. Rojas, and C. A. Davis (2017). The social media response to black lives matter: how twitter users interact with black lives matter through hashtag use. *Ethnic and Racial Studies 40*(11), 1814–1830.

Ito, K. and S. Zhang (2020). Willingness to pay for clean air: Evidence from air purifier markets in china. *Journal of Political Economy 128*(5), 1627–1672.

Jain, A. K., S. R. Sahoo, and J. Kaubiyal (2021). Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems 7*(5), 2157–2177.

James, A. G. and B. Smith (2020, June). Geographic dispersion of economic shocks: Evidence from the fracking revolution: Comment. *American Economic Review 110*(6), 1905–13.

Janssen, A. and E. Parslow (2021). Pregnancy persistently reduces alcohol purchases: Causal evidence from scanner data. *Health Economics 30*(2), 231–247.

Jaramillo, P. and N. Z. Muller (2016). Air pollution emissions and damages from energy production in the u.s.: 2002–2011. *Energy Policy 90*, 202–211.

Jin, X., A. Fiore, K. F. Boersma, I. D. Smedt, and L. Valin (2020). Inferring changes in summertime surface ozone–nox–voc chemistry over u.s. urban areas from two decades of satellite and ground-based observations. *Environmental Science & Technology 54*(11), 6518–6529. PMID: 32348127.

Jin, X., A. M. Fiore, K. L. Civerolo, J. Bi, Y. Liu, A. van Donkelaar, R. V. Martin, M. Z. Al-Hamdan, Y. Zhang, T. Z. Insaf, M.-A. Kioumourtzoglou, M. Z. He, and P. L. Kinney (2019). Comparison of multiple pm 2.5 exposure products for estimating health benefits of emission controls over new york state, usa. *Environmental Research Letters*.

Johnson, T., S. Guttikunda, G. Wells, P. Artaxo, T. Bond, A. Russell, J. Watson, and J. West (2011, 01). *Tools for Improving Air Quality Management: A Review of Top-Down Source Apportionment Techniques and Their Application in Developing Countries.*

Jones, B. A. (2022). Dust storms and violent crime. *Journal of Environmental Economics and Management 111*, 102590.

Kaye, D., D. Mijatovic, E. Lanza, and P. Tlakula (2016). Joint declaration on freedom of expression and countering violent extremism.

Kendix, M. and W. Walls (2010). Estimating the impact of refinery outages on petroleum product prices. *Energy Economics 32*(6), 1291–1298.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics 89*(11), 2093–2119.

Kramer, A. D. I., J. E. Guillory, and J. T. Hancock (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences 111*(24), 8788–8790.

Kuchler, T., Y. Li, L. Peng, J. Stroebel, and D. Zhou (2020, June). Social proximity to capital: Implications for investors and firms. Working Paper 27299, National Bureau of Economic Research.

Latonero, M. and I. Shklovski (2011, 10). Emergency management, twitter, and social media evangelism. *IJISCRAM 3*, 1–16.

Lavaine, E. and M. Neidell (2017). Energy production and health externalities: Evidence from oil refinery strikes in france. *Journal of the Association of Environmental and Resource Economists 4*(2), 447–477.

Ledgerwood, D. and N. Petry (2015, 12). Pathological gambling and gambling disorder. *International Encyclopedia of the Social & Behavioral Sciences*.

Lee, D. S., J. McCrary, M. J. Moreira, and J. R. Porter (2021, August). Valid t-ratio inference for iv. Working Paper 29124, National Bureau of Economic Research.

Levesque, S., T. Taetzsch, M. E. Lull, U. Kodavanti, K. Stadler, A. Wagner, J. A. Johnson, L. Duke, P. Kodavanti, M. J. Surace, et al. (2011). Diesel exhaust activates and primes microglia: air pollution, neuroinflammation, and regulation of dopaminergic neurotoxicity. *Environmental health perspectives 119*(8), 1149–1155.

Levinson, A. (2012). Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics 96*(9), 869–880.

Linden, A. (2019, April). RETRODESIGN: Stata module to compute type-S (Sign) and type-M (Magnitude) errors. Statistical Software Components, Boston College Department of Economics.

Ljoså, C. H. and B. Lau (2009). Shiftwork in the norwegian petroleum industry: overcoming difficulties with family and social life–a cross sectional study. *Journal of Occupational Medicine and Toxicology 4*(1), 1–10.

Loughery, J., T. Arciszewski, K. Kidd, A. Mercer, L. M. Hewitt, D. Maclatchy, and K. Munkittrick (2013, 11). Understanding the chronic impacts of oil refinery wastewater requires consideration of sediment contributions to toxicity. *Archives of environmental contamination and toxicology 66*.

Lu, H.-P. and K.-L. Hsiao (2009, 02). Gender differences in reasons for frequent blog posting. *Online Information Review 33*, 135–156.

Lyons, C. (2007). Community (dis)organization and racially motivated crime. *American Journal of Sociology 113*(3), 815–863.

MacAvaney, S., H.-R. Yao, E. Yang, K. Russell, N. Goharian, and O. Frieder (2019, 08). Hate speech detection: Challenges and solutions. *PLOS ONE 14*(8), 1–16.

MacKerron, G. and S. Mourato (2009). Life satisfaction and air quality in london. *Ecological Economics 68*(5), 1441–1453.

Mahajan, R. and V. Mansotra (2021). Correlating crime and social media: Using semantic sentiment analysis. *International Journal of Advanced Computer Science and Applications 12*.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal 16*(1), S1–S23.

Martins, N. and A. Weaver (2019). The role of media exposure on relational aggression: A meta-analysis. *Aggression and Violent Behavior 47*, 90–99.

McCrary, J. and S. Sanga (2021, 05). The Impact of the Coronavirus Lockdown on Domestic Violence. *American Law and Economics Review 23*(1), 137–163.

Mellgren, C., M. Andersson, and A.-K. Ivert (2017). For whom does hate crime hurt more? a comparison of consequences of victimization across motives and crime types. *Journal of Interpersonal Violence 36*, NP1512 – 1536NP.

Merchant, R. M. and N. Lurie (2020, 05). Social Media and Emergency Preparedness in Response to Novel Coronavirus. *JAMA 323*(20), 2011–2012.

Miller, T. R., M. A. Cohen, D. I. Swedler, B. Ali, and D. V. Hendrie (2021). Incidence and costs of personal and property crimes in the usa, 2017. *Journal of Benefit-Cost Analysis 12*(1), 24–54.

Millet, D. B., D. J. Jacob, K. F. Boersma, T.-M. Fu, T. P. Kurosu, K. Chance, C. L. Heald, and A. Guenther (2008). Spatial distribution of isoprene emissions from north america derived from formaldehyde column measurements by the omi satellite sensor. *Journal of Geophysical Research: Atmospheres 113*(D2).

Mitts, T. (2019). From isolation to radicalization: Anti-muslim hostility and support for isis in the west. *American Political Science Review 113*(1), 173–194.

Mitts, T., G. Phillips, and B. F. Walter (2022). Studying the impact of isis propaganda campaigns. *The Journal of Politics 84*(2), 1220–1225.

Moretti, E. and M. Neidell (2011). Pollution, health, and avoidance behavior: Evidence from the ports of los angeles. *The Journal of Human Resources 46*(1), 154–175.

Mu, Y., E. A. Rubin, and E. Zou (2021, April). What's missing in environmental (self-)monitoring: Evidence from strategic shutdowns of pollution monitors. Working Paper 28735, National Bureau of Economic Research.

Muller, K. and C. Schwarz (2020a, 10). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association 19*(4), 2131–2167.

Muller, K. and C. Schwarz (2020b, 7). From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment. *Available at SSRN*.

Muller, N. Z. and R. Mendelsohn (2009). Efficient pollution regulation: Getting the prices right. *The American Economic Review 99*(5), 1714–1739.

Muntifering, M. (2021, March). Air pollution, investor sentiment and excessive returns. *Journal of Asset Management 22*(2), 110–119.

Murphy, S. R., E. S. Schelegle, L. A. Miller, D. M. Hyde, and L. S. Van Winkle (2013). Ozone exposure alters serotonin and serotonin receptor expression in the developing lung. *toxicological sciences 134*(1), 168–179.

Musi, B., G. Dell'Omo, L. Ricceri, D. Santucci, G. Laviola, G. Bignami, and E. Alleva (1994). Effects of acute and continuous ozone (o3) exposure on activity/exploration and social behavior of cd-1 mice. *Neurotoxicology 15*(4), 827—835.

Mustaqim, T., K. Umam, and M. A. Muslim (2020, jun). Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm. *Journal of Physics: Conference Series 1567*, 032024.

Myers, W., T. McSween, R. Medina, K. Rost, and A. Alvero (2010, 10). The implementation and maintenance of a behavioral safety process in a petroleum refinery. *Journal of Organizational Behavior Management 30*, 285–307.

Nakazato, R. K., M. C. Rinaldi, and M. Domingos (2015). Will technological modernization for power generation at an oil refinery diminish the risks from air pollution to the atlantic rainforest in cubatão, se brazil? *Environmental Pollution 196*, 489–496.

Nattero, G. and A. Enrico (1996). Outdoor pollution and headache. *Headache: The Journal of Head and Face Pain 36*(4), 243–245.

Nauze, A. L. and E. R. Severnini (2021, May). Air pollution and adult cognition: Evidence from brain training. Working Paper 28785, National Bureau of Economic Research.

Newman, H. and D. Joyner (2018). Sentiment analysis of student evaluations of teaching. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, and B. du Boulay (Eds.), *Artificial Intelligence in Education*, Cham, pp. 246–250. Springer International Publishing.

Ng, S., B. Hollingsworth, E. Busey, J. L. Wandell, D. Miles, and J. Poti (2018). Federal nutrition program revisions impact low-income households' food purchases. *American journal of preventive medicine 54 3*, 403–412.

Nguyen, G., H. Nguyen, and H. Pham (2021, September). The Role of Social Connectedness: Evidence from Mergers and Acquisitions. Working Papers 2116, Waseda University, Faculty of Political Science and Economics.

Nijhawan, L. P., M. D. Janodia, B. Muddukrishna, K. M. Bhat, K. L. Bairy, N. Udupa, P. B. Musmade, et al. (2013). Informed consent: Issues and challenges. *Journal of advanced pharmaceutical technology & research 4*(3), 134.

Nimubona, A.-D. and B. Sinclair-Desgagne (2013). The Pigouvian Tax Rule in the Presence of an Eco-Industry. *Economics Bulletin 33*(1), 747–752.

Obradovich, N., D. Tingley, and I. Rahwan (2018). Effects of environmental stressors on daily governance. *Proceedings of the National Academy of Sciences 115*(35), 8710–8715.

Oladimeji, T., J. Sonibare, M. Odunfa, and O. Obanla (2015, 02). Environmental impact analysis of the emission from petroleum refineries in nigeria. *Energy and Environment Research 5*.

Ozdamar, O. and E. Giovanis (2017). The causal effects of income support and housing benefits on mental well-being: An application of a bayesian network. *Metroeconomica 68*(3), 398–424.

Paluck, E. (2009, 04). Reducing intergroup prejudice and conflict using the media: A field experiment in rwanda. *Journal of personality and social psychology 96*, 574–87.

Panger, G. (2016). Reassessing the facebook experiment: critical thinking about the validity of big data research. *Information, Communication & Society 19*(8), 1108–1126.

Parrish, D. D., T. B. Ryerson, J. Mellqvist, J. Johansson, A. Fried, D. Richter, J. G. Walega, R. A. Washenfelder, J. A. de Gouw, J. Peischl, K. C. Aikin, S. A. McKeen, G. J. Frost, F. C. Fehsenfeld, and S. C. Herndon (2012). Primary and secondary sources of formaldehyde in urban atmospheres: Houston texas region. *Atmospheric Chemistry and Physics 12*(7), 3273–3288.

Patrikarakos, D. (2017). *War in 140 Characters: How Social Media is Reshaping Conflict in the Twenty-First Century*. NY: Basic Books.

Pennington, J., R. Socher, and C. D. Manning (2014). Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.

Persico, C. L. and K. R. Johnson (2021). The effects of increased pollution on covid-19 cases and deaths. *Journal of environmental economics and management 107*, 102431.

Petruzzi, S., M. Fiore, G. Dell'Omo, and E. Alleva (1995). Exposure to ozone inhibits isolation-induced aggressive behavior of adult cd-1 male mice. *Aggressive Behavior 21*(5), 387–396.

Pezzella, F. S., M. D. Fetzer, and T. Keller (2019). The dark figure of hate crime underreporting. *American Behavioral Scientist 0*(0), 0002764218823844.

Pierson, E., C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, and S. Goel (2020, 07). A large-scale analysis of racial disparities in police stops across the united states. *Nature Human Behaviour 4*, 1–10.

Pigou, A. (1920, 01). *The Economics of Welfare*.

Ragothaman, A. and W. A. Anderson (2017). Air quality impacts of petroleum refining and petrochemical industries. *Environments 4*(3).

Rajper, S. A., S. Ullah, and Z. Li (2018, 03). Exposure to air pollution and self-reported effects on chinese students: A case study of 13 megacities. *PLOS ONE 13*(3), 1–21.

Ranson, M. (2014). Crime, weather, and climate change. *Journal of environmental economics and management 67*(3), 274–302.

Reardon, S. F., S. A. Matthews, D. O'sullivan, B. A. Lee, G. Firebaugh, C. R. Farrell, and K. Bischoff (2008). The geographic scale of metropolitan racial segregation. *Demography 45*(3), 489–514.

Rehbein, O. and S. Rother (2020, October). The Role of Social Networks in Bank Lending. CRC TR 224 Discussion Paper Series crctr224_2020_162v2, University of Bonn and University of Mannheim, Germany.

Rehdanz, K. and D. Maddison (2008). Local environmental quality and life-satisfaction in germany. *Ecological Economics 64*(4), 787–797.

Rosenquist, J. N., J. H. Fowler, and N. A. Christakis (2011). Social network determinants of depression. *Molecular Psychiatry 16*, 273–281.

Rotton, J. (1983). Affective and cognitive consequences of malodorous pollution. *Basic and applied social psychology 4*(2), 171–191.

Rotton, J., T. Barry, J. Frey, and E. Soler (1978). Air pollution and interpersonal attraction 1. *Journal of Applied Social Psychology 8*(1), 57–71.

Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics 6*(1), 34 – 58.

Saez, E., J. Slemrod, and S. H. Giertz (2012, March). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

Sandagiri, C., B. T. Kumara, and B. Kuhaneswaran (2021). Deep neural network-based crime prediction using twitter data. *International Journal of Systems and Service-Oriented Engineering (IJSSOE) 11*(1), 15–30.

Schlenker, W. and M. J. Roberts (2009). Nonlinear temperature effects indicate severe damages to u.s. crop yields under climate change. *Proceedings of the National Academy of Sciences 106*(37), 15594–15598.

Schlenker, W. and W. R. Walker (2015, 10). Airports, Air Pollution, and Contemporaneous Health. *The Review of Economic Studies 83*(2), 768–809.

Schumann, S. and Y. Moore (2022). The covid-19 outbreak as a trigger event for sinophobic hate crimes in the united kingdom. *The British Journal of Criminology*.

Selinger, C., M. Choisy, and S. Alizon (2021). Predicting covid-19 incidence in french hospitals using human contact network analytics. *International Journal of Infectious Diseases 111*, 100–107.

Shadbegian, R. J. and W. B. Gray (2005). Pollution abatement expenditures and plant-level productivity: A production function approach. *Ecological Economics 54*(2), 196–208. Technological Change and the Environment.

Sharma, S. (2013, 01). Environmental regulation, abatement, and productivity: A frontier analysis. *SSRN Electronic Journal*.

Shelar, A. and C.-Y. Huang (2018). Sentiment analysis of twitter data. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1301–1302.

Shepherd, H., F. Atherden, H. Chan, A. Loveridge, and A. Tatem (2021, 12). Domestic and international mobility trends in the united kingdom during the covid-19 pandemic: an analysis of facebook data. *International Journal of Health Geographics 20*.

Siriaraya, P., Y. Zhang, Y. Wang, Y. Kawai, M. Mittal, P. Jeszenszky, and A. Jatowt (2019). Witnessing crime through tweets: a crime investigation tool based on social media. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 568–571.

Soulage, C., D. Perrin, J.-M. Cottet-Emard, J. Pequignot, Y. Dalmaz, and J.-M. Pequignot (2004). Central and peripheral changes in catecholamine biosynthesis and turnover in rats after a short period of ozone exposure. *Neurochemistry international 45*(7), 979–986.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stavins, R. and L. Bennear (2007, 02). Second-best theory and the use of multiple policy instruments. *Environmental & Resource Economics 37*, 111–129.

Stein, R. A., P. A. Jaques, and J. F. Valiati (2019). An analysis of hierarchical text classification using word embeddings. *Information Sciences 471*, 216–232.

Stella, M., E. Ferrara, and M. D. Domenico (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences 115*(49), 12435–12440.

Sweeney, R. (2015). Environmental regulation, imperfect competition and market spillovers: The impact of the 1990 clean air act amendments on the us oil refining industry.

Tao, Y., F. Zhang, C. Shi, and Y. Chen (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. *Sustainability 11*(18).

Tausczik, Y. R. and J. W. Pennebaker (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology 29*(1), 24–54.

Taylor, P. J. (1967). Shift and day work: A comparison of sickness absence, lateness, and other absence behaviour at an oil refinery from 1962 to 1965. *British Journal of Industrial Medicine 24*(2), 93–102.

Thomas, Q. (2015). Social media's influence on public discourse in the pacific northwest.

Tice, D., E. Bratslavsky, and R. Baumeister (2001, 02). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it! *Journal of personality and social psychology 80*, 53–67.

Uboh, F., M. Akpanabiatu, I. Ekaidem, P. Ebong, and I. Umoh (2007). Effect of inhalation exposure to gasoline on sex hormones profile in wistar albino rats. *ACTA ENDOCRINOLOGICA-BUCHAREST- 3*(1), 23.

UN (2019). The united nations strategy and plan of action on hate speech.

Van Berlo, D., C. Albrecht, A. M. Knaapen, F. R. Cassee, M. E. Gerlofs-Nijland, I. M. Kooter, N. Palomero-Gallagher, H.-J. Bidmon, F.-J. van Schooten, J. Krutmann, et al. (2010). Comparative evaluation of the effects of short-term inhalation exposure to diesel engine exhaust on rat lung and brain. *Archives of toxicology 84*(7), 553–562.

Vicario, M. D., A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences 113*(3), 554–559.

Voon, V., A. Mehta, and M. Hallett (2011, 06). Impulse control disorders in parkinson's disease: Recent advances. *Current opinion in neurology 24*, 324–30.

Wachasunder, S. (2004). Assesment of refinery noise impact on workers – a case study. *International Journal of Environmental Studies 61*(4), 459–470.

Wang, J., Z. Wang, D. Zhang, and J. Yan (2017). Combining knowledge with deep convolutional neural networks for short text classification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 2915–2921.

Wang, L., Y. Zheng, and S. C. Buck (2021). How does the affordable care act medicaid expansion affect cigarette consumption?—the mechanism and heterogeneity. *Journal of Consumer Affairs*.

Wang, T. (2021a, September). Media, pulpit, and populist persuasion: Evidence from father coughlin. *American Economic Review 111*(9), 3064–92.

Wang, T. (2021b). Waves of empowerment: Black radio and the civil rights movement. *Working paper*.

Wang, Y., W. Yu, S. Liu, and S. D. Young (2019). The relationship between social media data and crime rates in the united states. *Social Media + Society 5*(1), 2056305119834585.

Welsch, H. (2006). Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics 58*(4), 801–813.

West, S. E. and R. C. Williams (2004, March). Empirical estimates for environmental policy making in a second-best setting. Working Paper 10330, National Bureau of Economic Research.

Wickramasekera, N., J. Wright, H. Elsey, J. Murray, and S. Tubeuf (2015). Cost of crime: A systematic review. *Journal of Criminal Justice 43*(3), 218–228.

Williams, M. L., P. Burnap, A. Javed, H. Liu, and S. Ozalp (2019, 07). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology 60*(1), 93–117.

Williams, M. L., P. Burnap, A. Javed, H. Liu, and S. Ozalp (2020). Hate in the machine: Anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology 60*(1), 93–117.

Wood, W., F. Wong, and J. Chachere (1991, 06). Effects of media violence on viewers' aggression in unconstrained social interaction. *Psychological bulletin 109*, 371–83.

Wright, A., H. Rickards, and A. E. Cavanna (2012). Impulse-control disorders in gilles de la tourette syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences 24*(1), 16–27. PMID: 22450610.

Xu, A., C. Liu, Y. Wan, Y. Bai, and Z. Li (2021). Monkeys fight more in polluted air. *Scientific Reports 11*(1), 1–7.

Yang, W., S. Kandula, M. Huynh, S. Greene, G. Wye, W. Li, H. Chan, E. McGibbon, A. Yeung, D. Olson, A. Fine, and J. Shaman (2020, 10). Estimating the infection-fatality risk of sars-cov-2 in new york city during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases 21*.

Yen, Y.-S. (2016, 11). Factors enhancing the posting of negative behavior in social media and its impact on venting negative emotions. *Management Decision 54*, 2462–2484.

Zhang, X., X. Chen, and X. Zhang (2018). The impact of exposure to air pollution on cognitive performance. *Proceedings of the National Academy of Sciences 115*(37), 9193–9197.

Zhang, X., X. Zhang, and X. Chen (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being? *Journal of Environmental Economics and Management 85*, 81–94.

Zheng, S., J. Wang, C. Sun, X. Zhang, and M. Kahn (2019, 03). Air pollution lowers chinese urbanites' expressed happiness on social media. *Nature Human Behaviour 3*.

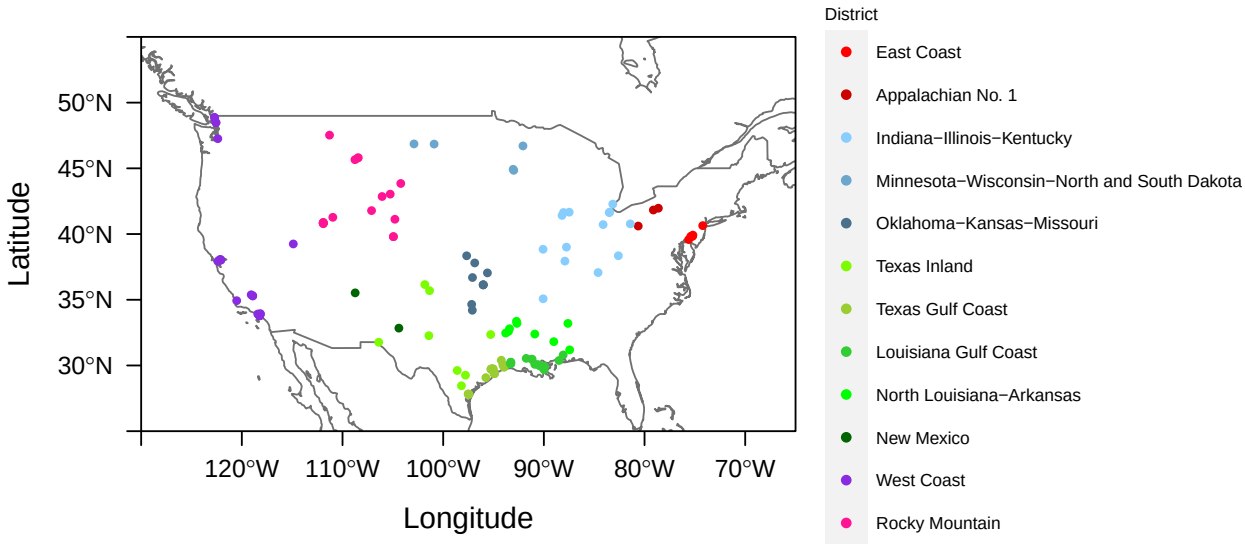
Zhong, N., J. Cao, and Y. Wang (2017). Traffic congestion, ambient air pollution, and health: Evidence from driving restrictions in beijing. *Journal of the Association of Environmental and Resource Economists 4*(3), 821 – 856.

Zhu, L., D. Jacob, L. Mickley, E. Marais, D. Cohan, Y. Yoshida, B. Duncan, G. G. Abad, and K. Chance (2014). Anthropogenic emissions of highly reactive volatile organic compounds in eastern texas inferred from oversampling of satellite (omi) measurements of hcho columns. *Environmental Research Letters 9*, 114004.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*(1), 415–438.

Zou, E. Y. (2021, July). Unwatched pollution: The effect of intermittent monitoring on air quality. *American Economic Review 111*(7), 2101–26.

Figure 1: Locations of operating refinery plants in different districts



Notes: Locations of operating refinery plants are obtained from the EIA and Refinitiv. There are 135 operating plants in 12 districts. They are widely distributed across the country, with higher plant concentrations in Texas, Louisiana, California, and Oklahoma.

Table 1: Summary statistics of surrounding HCHO on outage days and operating days

	HCHO
Normal operation	8.137
	[5.594]
	(100,773)
Planned day1 (*PlannedShut*)	7.638
	[5.227]
	(174)
Planned day2+ (*PlannedDowntime*)	7.680
	[4.967]
	(2,093)
Unplanned day1 (*UnplannedShut*)	9.122
	[6.166]
	(295)
Unplanned day2+ (*UnplannedDowntime*)	8.132
	[5.416]
	(2,000)
All days	8.129
	[5.581]
	(105,335)

Notes: This table reports the mean, standard deviation, and observations of HCHO at the plant-day level with and without outages. The sample includes daily observations within 20km of 101 refinery plants 2014-2019, with some random missing in satellite reports. The unit of HCHO column density is 10^{15} molecules/cm^2. Standard deviations are reported in brackets. The number of days is reported under the brackets.

Table 2: Effects of temporary outages on surrounding HCHO

	HCHO			
	Panel A: Full controls			
	(1)	(2)	(3)	(4)
PlannedShut	-0.013	-0.089	-0.016	-0.015
	(0.227)	(0.221)	(0.227)	(0.228)
PlannedDowntime	-0.449***	-0.492***	-0.450***	-0.451***
	(0.079)	(0.075)	(0.079)	(0.079)
UnplannedShut	0.599**	0.613**	0.598**	0.599**
	(0.248)	(0.253)	(0.248)	(0.248)
UnplannedDowntime	-0.192*	-0.176	-0.194*	-0.195*
	(0.107)	(0.106)	(0.107)	(0.107)
OutageAnyPlant	.0302	.00112	.0119	-.000804
	(.105)	(.171)	(.105)	(.108)
Days			-6.18**	-5.9*
			(2.85)	(3.14)
$Days^2$				-.142
				(.216)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
	Panel B: Short controls			
UnplannedShut	0.605**	0.615**	0.604**	0.605**
	(0.246)	(0.251)	(0.246)	(0.246)
Downtime	-0.352***	-0.373***	-0.354***	-0.355***
	(0.054)	(0.050)	(0.054)	(0.054)
OutageAnyPlant	0.032	0.002	0.013	0.001
	(0.105)	(0.171)	(0.104)	(0.108)
Days			-6.192**	-5.914*
			(2.846)	(3.129)
$Days^2$				-.141
				(.214)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y

Notes: Panel A reports estimated coefficients from equation (1). The sample is daily HCHO levels within 20km of 101 refineries 2014-2019. Given the negative estimates on *PlannedShut*, *PlannedDowntime* and *UnplannedDowntime*, I combine these three variables and estimate a short equation in Panel B. All regressions include plant fixed effects, day of week (DOW) fixed effects, and single time term *OutageAnyPlant*. Column (1) includes year and month fixed effects. Based on Column (1), linear and quadratic day trends are added in Column (3) and (4), respectively. Column (2) includes year by month fixed effects and shows a same-month comparison. The mean and standard deviation of dependent variable is reported at the bottom of each panel, and the unit is 10^{15} molecules/cm^2. Standard errors are clustered at the plant level.

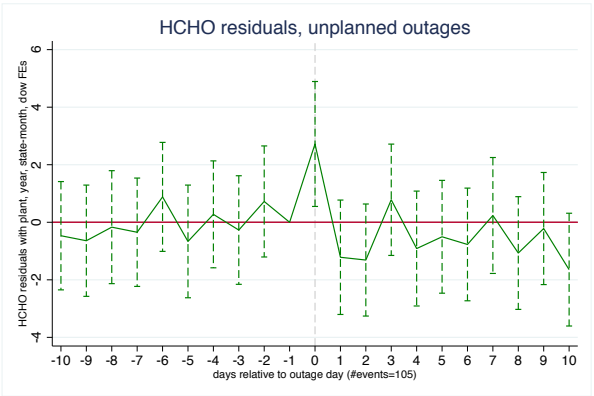
Table 3: Effects of unplanned shutdowns on tweets

	Panel A: HCHO			
	(1)	(2)	(3)	(4)
UnplannedShut	5.263***	5.223***	5.259***	5.259***
	(0.875)	(0.877)	(0.875)	(0.875)
Downtime	-0.385	-0.394	-0.387	-0.392
	(0.248)	(0.252)	(0.248)	(0.248)
Observations	10275	10275	10275	10275
R-square	0.139	0.145	0.139	0.139
Y-mean	8.452	8.452	8.452	8.452
Y-sd	5.701	5.701	5.701	5.701
	Panel B: Proportion of tweets with air pollution keywords (in percentage)			
UnplannedShut	0.457**	0.460**	0.460**	0.462**
	(0.198)	(0.198)	(0.198)	(0.198)
Downtime	-0.080	-0.046	-0.078	-0.083
	(0.060)	(0.061)	(0.060)	(0.060)
Observations	19361	19361	19361	19361
R-square	0.015	0.023	0.016	0.016
Y-mean	0.564	0.564	0.564	0.564
Y-sd	1.800	1.800	1.800	1.800
	Panel C: Proportion of tweets with health keywords (in percentage)			
UnplannedShut	0.002	0.011	0.001	0.001
	(0.047)	(0.047)	(0.047)	(0.047)
Downtime	-0.002	0.002	-0.003	-0.003
	(0.014)	(0.014)	(0.014)	(0.014)
Observations	19361	19361	19361	19361
R-square	0.009	0.013	0.009	0.009
Y-mean	0.065	0.065	0.065	0.065
Y-sd	0.423	0.423	0.423	0.423
	Panel D: Sentiment			
UnplannedShut	-0.001	-0.001	-0.001	-0.001
	(0.010)	(0.010)	(0.010)	(0.010)
Downtime	-0.003	-0.000	-0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	21233	21233	21233	21233
R-square	0.236	0.268	0.236	0.242
Y-mean	0.122	0.122	0.122	0.122
Y-sd	0.108	0.108	0.108	0.108
	Panel E: #Offensive tweets			
UnplannedShut	29.963*	35.300**	29.949*	33.560**
	(13.060)	(11.528)	(13.020)	(10.902)
Downtime	25.426	14.487	25.393	34.554
	(51.805)	(45.725)	(51.732)	(47.202)
Observations	10275	10275	10275	10275
R-square	0.477	0.507	0.477	0.488
Y-mean	137.592	137.592	137.592	137.592
Y-sd	304.289	304.289	304.289	304.289

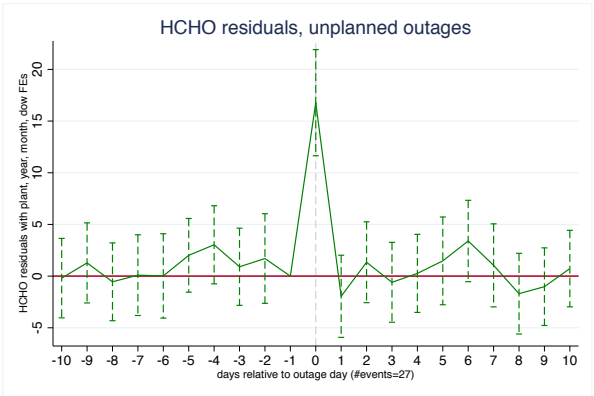
	Panel F: #Racist tweets			
UnplannedShut	5.089*	5.830*	5.087*	5.615**
	(2.445)	(2.404)	(2.441)	(2.156)
Downtime	5.507	3.381	5.504	6.844
	(9.309)	(8.323)	(9.300)	(8.692)
Observations	10275	10275	10275	10275
R-square	0.422	0.450	0.422	0.430
Y-mean	18.647	18.647	18.647	18.647
Y-sd	50.834	50.834	50.834	50.834
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: This table shows reduced form results by estimating a short form of equation (1). The sample is daily tweet outcomes within 20km of the top ten refinery plants (specified in Table S7.1) 2014-2019. All regressions include plant fixed effects, day of week (DOW) fixed effects, and single time term *OutageAnyPlant*. Column (1) includes year and month fixed effects. Based on Column (1), linear and quadratic day trends are added in Column (3) and (4), respectively. Column (2) includes year by month fixed effects and shows a same-month comparison. The mean and standard deviation of dependent variable is reported at the bottom of each panel. Standard errors are clustered at the plant level.

Figure 2: Effects of unplanned shutdowns on surrounding HCHO



(a) Abnormal outages



(b) Abnormal outages, top ten plants

Notes: These figures display the treatment effect of refinery plants' temporary unplanned outages on surrounding HCHO levels in each day relative to the pre-treatment period (day -1). In Panel (a), I use big plants with capacity above 200,000 barrels per day. In Panel (b), I use top ten refinery plants in Table S7.1. Both panels only use balanced events without any missing satellite reports in the 11-day event window. Figures without plant capacity constraints are shown in Figure S3.5. The green solid lines display the estimated coefficients after controlling for plant, year, state-month, and day of week fixed effects. The green dash bars show 95% confidence intervals. The number of events is reported in parentheses.

Table 4: Effects of unplanned shutdowns on hate crimes

	Panel A: #Hate crime events against black people			
	(1)	(2)	(3)	(4)
UnplannedShut	0.018**	0.018**	0.018**	0.017**
	(0.008)	(0.009)	(0.008)	(0.008)
Downtime	-0.002**	-0.002	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	105335	105335	105335	105335
R-square	0.151	0.154	0.151	0.151
Y-mean	0.031	0.031	0.031	0.031
Y-sd	0.192	0.192	0.192	0.192
	Panel B: #Black victims			
UnplannedShut	0.025*	0.027*	0.025*	0.025*
	(0.015)	(0.015)	(0.015)	(0.015)
Downtime	-0.004	-0.003	-0.004	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	105335	105335	105335	105335
R-square	0.087	0.090	0.087	0.087
Y-mean	0.032	0.032	0.032	0.032
Y-sd	0.277	0.277	0.277	0.277
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: This table shows reduced form results by estimating a short form of equation (1). The sample is daily hate crime outcomes in the county where refinery plants are located 2014-2019. Counties without any refinery are not included in the regression. *UnplannedShut* indicates there is at least one refinery plant in this county experiencing an unplanned outage, and this day is the first day of the shutdown. All regressions include county fixed effects, day of week (DOW) fixed effects, and single time term *OutageAnyPlant*. Column (1) includes year and month fixed effects. Based on Column (1), linear and quadratic day trends are added in Column (3) and (4), respectively. Column (2) includes year by month fixed effects and shows a same-month comparison. The mean and standard deviation of dependent variable is reported at the bottom of each panel. Standard errors are clustered at the county level.

Table 5: Effects of unplanned shutdowns on other crimes

	Panel A: #Crimes			
	(1)	(2)	(3)	(4)
UnplannedShut	5.094**	5.677*	5.087**	5.179**
	(2.308)	(2.911)	(2.318)	(2.289)
Downtime	-0.622	-0.952	-0.642	-0.593
	(1.163)	(1.065)	(1.154)	(1.016)
Observations	61831	61831	61831	61831
R-square	0.557	0.562	0.557	0.558
Y-mean	42.963	42.963	42.963	42.963
Y-sd	107.560	107.560	107.560	107.560
	Panel B: #Victims			
UnplannedShut	6.435**	7.191*	6.428**	6.540**
	(2.783)	(3.547)	(2.794)	(2.759)
Downtime	-1.081	-1.516	-1.105	-1.044
	(1.513)	(1.313)	(1.502)	(1.341)
Observations	61831	61831	61831	61831
R-square	0.524	0.530	0.525	0.525
Y-mean	49.906	49.906	49.906	49.906
Y-sd	128.149	128.149	128.149	128.149
County FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
Trends			Linear	Quadratic

Notes: This table shows reduced form results by estimating a short form of equation (1). The sample is daily non-hate crime outcomes in the county where refinery plants are located 2014-2019. Counties without any refinery or without any reported crimes 2014-2019 are not included in the regression. Given a lower coverage of non-hate crime reporting system, sample sizes in this table are smaller than those in Table 4. *UnplannedShut* indicates there is at least one refinery plant in this county experiencing an unplanned outage, and this day is the first day of the shutdown. All regressions include county fixed effects, day of week (DOW) fixed effects, and single time term $OutageAnyPlant$. Column (1) includes year and month fixed effects. Based on Column (1), linear and quadratic day trends are added in Column (3) and (4), respectively. Column (2) includes year by month fixed effects and shows a same-month comparison. The mean and standard deviation of dependent variable is reported at the bottom of each panel. Standard errors are clustered at the county level.

Table 6: Effects of unplanned shutdowns on benzene by monitor distances

	Benzene from ground monitors				
	(1)	(2)	(3)	(4)	(5)
	1km	3km	5km	7km	9km
UnplannedShut	1.120*	0.691*	0.244**	0.210***	0.172***
	(0.486)	(0.298)	(0.102)	(0.035)	(0.029)
Downtime	-0.458	-0.331	-0.268*	-0.062	-0.108*
	(0.274)	(0.290)	(0.121)	(0.044)	(0.051)
Observations	1029	1651	4017	15647	17884
R-square	0.198	0.204	0.205	0.158	0.167
Y-mean	1.564	1.518	1.383	1.274	1.259
Y-sd	1.851	1.835	1.352	1.604	1.537
	15km	20km	25km	30km	35km
UnplannedShut	0.151***	0.144***	0.132***	0.110***	0.101***
	(0.035)	(0.027)	(0.030)	(0.022)	(0.023)
Downtime	-0.087**	-0.054	-0.055*	-0.030	-0.026
	(0.036)	(0.035)	(0.030)	(0.041)	(0.036)
Observations	22924	25284	27223	36781	40269
R-square	0.170	0.185	0.192	0.160	0.167
Y-mean	1.356	1.318	1.321	1.282	1.245
Y-sd	1.331	1.176	1.150	1.203	1.155
Plant FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

Notes: This table shows first stage results by estimating a short form of equation (1). I conduct a plant-day level analysis and use benzene monitor reports as the dependent variable. In the upper panel Column (1), for each plant, I calculate average benzene reports on that day if monitors are located within 1km of that plant. In upper Column (2)-(5), I use larger radiuses of 3km to 9km and even larger distance cutoffs in the bottom panel. As distance increases, sample size becomes larger, and estimation precision improves. All regressions include plant fixed effects, day of week (DOW), year and month fixed effects, and quadratic day trends. The mean and standard deviation of the dependent variable is reported at the bottom of each panel, and the unit is parts per billion carbon. Standard errors are clustered at the plant level.

Table 7: Effects of unplanned shutdowns on medical expenditure

	Expenditure at the county-day level ($)		
	Cough remedies	Sinus remedies	Breathing aids
(Product module code)	(8425)	(8502)	(7790)
UnplannedShut	0.652***	0.068**	0.030***
	(0.250)	(0.033)	(0.011)
Downtime	-0.028	-0.007	-0.014
	(0.071)	(0.009)	(0.011)
Observations	105335	105335	105335
R-square	0.155	0.012	0.023
Y-mean	1.113	0.179	0.031
Y-sd	4.746	2.090	0.663
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: This table shows reduced form results by estimating a short form of equation (1). The sample is each household's daily average consumption expenditure in the county where refinery plants are located 2014-2019. Counties without any refinery are not included in the regression. *UnplannedShut* indicates there is at least one refinery plant in this county experiencing an unplanned outage, and this day is the first day of the shutdown. All regressions include county, year, month, day of week (DOW) fixed effects, quadratic time trends, and single time term *OutageAnyPlant*. Product module codes used for each dependent variable are reported in parentheses above the estimates. Detailed code list is available from Nielsen consumer panel (https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen). The mean and standard deviation of dependent variable is reported at the bottom, and the unit is dollar. Standard errors are clustered at the county level.

Table 8: Effects of unplanned shutdowns on foot traffic

	#Devices at the county-day level ($\times 10^5$)	#Visits at the county-day level ($\times 10^5$)	
		Amusement parks & recreational camps	General medical & surgical hospitals
	(1)	(2)	(3)
UnplannedShut	0.024	-0.067	2.488**
	(0.116)	(0.072)	(1.078)
Downtime	-0.168***	-0.065***	-0.592*
	(0.035)	(0.022)	(0.328)
Observations	30425	30425	30425
R-square	0.966	0.574	0.869
Y-mean	3.075	0.350	11.942
Y-sd	5.602	0.599	26.452
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: This table shows reduced form results by estimating a short form of equation (1). The sample is the daily number of active devices or visits in the county where refinery plants are located 2018-2019. Counties without any refinery are not included in the regression. As foot traffic data is only available after 2018, my study period of this analysis is 2018-2019, and I replicate the first stage results for 2018-2019 in Table S8.8. *UnplannedShut* indicates there is at least one refinery plant in this county experiencing an unplanned outage, and this day is the first day of the shutdown. All regressions include county, year, month, day of week (DOW) fixed effects, quadratic time trends, and single time term *OutageAnyPlant*. The mean and standard deviation of dependent variable is reported at the bottom, and the unit is the number of devices or visits divided by 10^5. Standard errors are clustered at the county level.

Table 9: Effects of unplanned shutdowns on mortality

	Panel A: #Death at the county--day level ($\times 10^{-3}$)			
	(1)	(2)	(3)	(4)
UnplannedShut	0.031	0.062	0.030	0.027
	(0.033)	(0.082)	(0.031)	(0.029)
Downtime	-0.047	-0.037	-0.051	-0.047
	(0.053)	(0.055)	(0.056)	(0.053)
Observations	105335	105335	105335	105335
R-square	0.075	0.076	0.075	0.076
Y-mean	0.911	0.911	0.911	0.911
Y-sd	33.458	33.458	33.458	33.458
	Panel B: Death proportion of sampled beneficiaries (in percentage $\times 10^{-3}$)			
UnplannedShut	0.004	0.008	0.004	0.003
	(0.004)	(0.011)	(0.004)	(0.004)
Downtime	-0.006	-0.005	-0.007	-0.006
	(0.007)	(0.007)	(0.007)	(0.007)
Observations	105335	105335	105335	105335
R-square	0.075	0.076	0.075	0.075
Y-mean	0.118	0.118	0.118	0.118
Y-sd	4.341	4.341	4.341	4.341
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: This table shows reduced form results by estimating a short form of equation (1). The sample is daily mortality measures in the county where refinery plants are located 2014-2019. I use the number of deaths in Panel A, and the proportion of death over sampled beneficiaries in that year in Panel B. Counties without any refinery are not included in the regression. *UnplannedShut* indicates there is at least one refinery plant in this county experiencing an unplanned outage, and this day is the first day of the shut-down. All regressions include county fixed effects, day of week (DOW) fixed effects, and single time term *OutageAnyPlant*. Column (1) includes year and month fixed effects. Based on Column (1), linear and quadratic day trends are added in Column (3) and (4), respectively. Column (2) includes year by month fixed effects and shows a same-month comparison. The mean and standard deviation of dependent variable is reported at the bottom of each panel, and the unit is deaths or death proportion times 1000. Standard errors are clustered at the county level.

Table 10: Effects of unplanned shutdowns on geographically distant tweets

	Sentiment	#Tweets ($\times 10^{-2}$) with			
		Pollution	Health	Offensive	Racist content
	(1)	(2)	(3)	(4)	(5)
UnplannedShut × Connected	-0.001	-1.967	0.126	0.163***	0.022***
	(0.013)	(1.767)	(0.571)	(0.022)	(0.003)
Observations	641856	5002053	5002053	5002053	5002053
R-square	0.021	0.001	0.001	0.028	0.013
Y-mean	0.108	2.397	0.352	4.850	1.538
Y-sd	0.302	57.047	19.517	19.458	6.159
County FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Connected$_{ij}$	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max	Max
Distance cutoff	150	150	150	150	150

Notes: This table shows reduced form results by estimating equation (2). The sample is daily tweet outcomes in geographically distant counties 2014-2019. Counties located within 150km of any refinery or without any geocoded tweets are not included in the regression. *UnplannedShut* indicates a connected refinery starting an abnormal outage. *Connected* is a static cross-sectional measure at the county pair level. I use the maximum treatment measure of all connected refineries to code *UnplannedShut* × *Connected* in this table, and alternative aggregation schemes in Table S9.11. Besides, I use alternative distance cutoffs to find geographically distant counties in Table S9.12. All regressions include recipient county, year, month, day of week (DOW) fixed effects, quadratic time trends, and single time term *OutageAnyPlant*. The mean and standard deviation of dependent variable is reported at the bottom, and the units are the average score between -1 and 1 and the number of tweets times 100. Standard errors are clustered at the county level.

Table 11: Effects of unplanned shutdowns on geographically distant crimes

	#Hate crimes ($\times 10^{-4}$)			#Other crimes
	all victims	Black	Asian	(absolute number)
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	1.285	8.283	-1.777	0.319***
	(12.694)	(6.468)	(1.426)	(0.084)
Observations	2381617	2381617	2381617	3060827
R-square	0.161	0.054	0.007	0.082
Y-mean	83.599	24.857	1.558	7.126
Y-sd	1030.102	517.791	128.448	21.298
	Panel B: #Crimes on lead day 1			
UnplannedShut × Connected	2.952	4.255	-2.404	0.375***
	(14.598)	(6.273)	(1.478)	(0.086)
R-square	0.161	0.054	0.007	0.082
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
Distance cutoff	150	150	150	150

Notes: This table shows reduced form results by estimating equation (2). The sample is daily hate crime and non-hate crime outcomes in geographically distant counties 2014-2019. Counties located within 150km of any refinery or without any crime reports are not included in the regression. *UnplannedShut* indicates a connected refinery starting an abnormal outage. *Connected* is a static cross-sectional measure at the county pair level. I use the maximum treatment measure of all connected refineries to code $UnplannedShut \times Connected$. All regressions include recipient county, year, month, day of week (DOW) fixed effects, quadratic time trends, and single time term *OutageAnyPlant*. The mean and standard deviation of dependent variable is reported at the bottom. Panel A reports same-day effects, while outcomes on the following day are merged with outage schedules in Panel B. The units are the number of non-hate crimes in Column (4) and the number of hate crimes times 10^{-4} in Column (1)-(3). Standard errors are clustered at the county level.

Table 12: Effects of unplanned shutdowns on geographically distant tweets and crimes, using Twitter SCI

	#Tweets ($\times 10^{-2}$) with		#Other crimes
	Offensive	Racist content	(absolute number)
	(1)	(2)	(3)
UnplannedShut × #Followers	0.022***	0.003***	0.063***
($\times 10^{-2}$)	(0.004)	(0.001)	(0.017)
Observations	5002053	5002053	3060827
R-square	0.046	0.020	0.090
Y-mean	4.850	1.538	7.126
Y-sd	19.458	6.159	21.298
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic
Aggregation	Max	Max	Max
Distance cutoff	150	150	150

Notes: This table shows reduced form results by estimating equation (5). The sample in Column (1) and (2) is daily tweet outcomes in geographically distant counties 2014-2019. The sample in Column (3) is daily non-hate crime outcomes in geographically distant counties 2014-2019. Counties located within 150km of any refinery or without any geocoded tweets or without any crime reports are not included in the regression. *UnplannedShut* indicates a connected refinery starting an abnormal outage. *#Followers* is a static cross-sectional measure at the county pair level. I use the maximum treatment measure of all connected refineries to code *UnplannedShut* × *#Followers* in this table. All regressions include recipient county, year, month, day of week (DOW) fixed effects, quadratic time trends, and single time term *OutageAnyPlant*. The mean and standard deviation of dependent variable is reported at the bottom. The units are the number of tweets times 100 in Column (1) and (2) and the number of non-hate crimes in Column (3). Standard errors are clustered at the recipient county level.

Online Appendix

S1 Empathetic preference with multiple externalities

This section extends empathetic preference theory by (Heal, 2021). Social contact does not directly affect utility, but shows up as weights that consumers place on others' utilities in their welfare functions. I explore how people's care about others affects the optimal pollution tax.

S1.1 Model setup

Consider a representative consumer in county i purchases good of quantity x_i and gains utility of $U_i(x_i)$. Local pollution due to production is P_i and the marginal damage of pollution is d. Consumer's wage rate w_i, time endowment T, and non-labor income I_i are considered given. Each consumer's utility function and budget constraint are:

$$U_i = U_i(x_i) - d \cdot P_i$$

$$I_i + w_i \cdot T \geq p_i \cdot x_i$$

Each person's welfare depends on his own utility and the utilities of others. Consumer's welfare maximization could be expressed as:

$$\max_{x_i} \ W_i = \sum_{j=1}^{J} (\gamma_i^j + \eta_i^j) \cdot U_j$$

where γ_i^j measures how much consumers in county i care about those in county j. η_i^j measures how much consumers in county i hate those in county j. I assume $0 \leq \gamma_i^j \leq 1$, $-1 \leq \eta_i^j \leq 0$, and $\gamma_i^i = 1$, $\eta_i^i = 0$. In other words, those living in the same community care about others at the maximum level. Besides, consumer in county i's perception and weights on others only depends on i's local factors. In other words, $\forall k \neq i$, $\partial \gamma_i^j / \partial x_k = 0$, $\partial \eta_i^j / \partial x_k = 0$.

Suppose that there is only one perfectly competitive industry. The production cost and price are $c_i(x_i)$ and p_i. For each firm in county i, its profit maximization problem is:

$$\max_{x_i} \ \pi_i = p_i \cdot x_i - c_i(x_i)$$

In the private optimal condition, $U_i'(x_i) - p_i = 0$, $p_i - c_i'(x_i) = 0$. To reach Pareto efficiency, we maximize social welfare taking all consumers and producers together:

$$\max_{x_1, x_2, \ldots, x_J} \ W = \sum_{i=1}^{J} \{ \sum_{j=1}^{J} (\gamma_i^j + \eta_i^j) \cdot U_j + I_i + w_i \cdot T - c_i(x_i) \}$$

In the social optimal condition, each x_i satisfies the first-order condition:

$$\frac{\partial W}{\partial x_i} = U_i'(x_i) - d \cdot e - c_i'(x_i) + \underbrace{[U_i'(x_i) - d \cdot e] \cdot \sum_{j \neq i} (\gamma_j^i + \eta_j^i)}_{\text{others care about } i} + \underbrace{\sum_{j=1}^{J} [\frac{\partial \gamma_i^j}{\partial x_i} + \frac{\partial \eta_i^j}{\partial x_i}] \cdot U_j}_{i \text{ cares about others}} = 0$$

where the fourth term captures how $U_i(x_i)$ affects others' welfare, and the last term shows how U_j affects consumer i's welfare. Both of them result from externalities of social contact.

S1.2 Optimal pollution tax

If we introduce a pollution tax T, new private optimum generates: $U_i'(x_i) - T \cdot e - c_i'(x_i) = 0$. To reach social optimum, optimal pollution tax T should satisfy:

$$T^* = d - [\frac{U_i'(x_i)}{e} - d] \cdot \sum_{j \neq i} (\gamma_j^i + \eta_j^i) - \sum_{j=1}^{J} [\frac{\partial \gamma_i^j}{\partial x_i} + \frac{\partial \eta_i^j}{\partial x_i}] \cdot U_j$$

First, assuming there is no externality interactions, $\partial \gamma_i^j / \partial x_i = 0$, $\partial \eta_i^j / \partial x_i = 0$. Optimal pollution tax is $T^* = d - [\frac{U_i'(x_i)}{e} - d] \cdot \sum_{j \neq i} (\gamma_j^i + \eta_j^i)$, and depends on county i's popularity perceived by others. I define counties where $\sum_{j \neq i} (\gamma_j^i + \eta_j^i) = 0$ as neutral counties. They are cared by some counties and hated by others, and positive and negative weights cancel out. Pollution tax in neutral counties should be $T^* = d$, the marginal external cost of pollution in the first-best tax design. In unpopular counties where residents are not cared by others, the dominant sentiment towards these residents is negative. Optimal pollution tax satisfies $T^* > d$.

With externality interactions, pollution makes people hate more and care less about others, and lower weights are placed on others' welfare. We have $\partial \gamma_i^j / \partial x_i < 0$ or $\partial \eta_i^j / \partial x_i < 0$. In this condition, T^* is higher than that without interactions. We should play tougher emission controls due to the low cost of forgone $U_i(x_i)$.

S2 Literature review: Impact of air pollution on health

I use medical expenditure, foot traffic, and mortality as dependent variables to study health impacts of air pollution. These outcomes are studied in earlier papers under different settings.

As a most important and costly outcome, infant, child, and adult mortality increase with air pollution. One commonly-used instrument for air pollution is the wind direction. Anderson (2019) leverage the quasi-random variation in pollution levels generated by wind patterns near highways in Los Angeles 1999-2001. The authors find doubling the percentage of time spent downwind of a highway increases mortality among individuals 75 and older by 3.6-6.8%. Deryugina et al. (2019) apply the wind direction instrument to the Medicare beneficiaries in the whole U.S., and finds a $1\mu g/m^3$ increase in $PM_{2.5}$ causes 0.69 additional deaths per million elderly individuals. Another instrument for air pollution is atmospheric inversion. An inversion occurs when a mass of hot air gets caught above a mass of cold air, trapping pollutants. Conditional on temperature, inversions themselves do not represent a health risk *per se* other than the accumulation of pollutants (Arceo et al., 2016). As such, Arceo et al. (2016) show an increase of 1ppb in CO and $1\mu g/m^3$ in PM_{10} instrumented by the atmospheric inversion results in 0.032 and 2.4 infant deaths per million new births respectively. Besides, some papers take advantage of environmental regulations, starts or suspensions of polluting events to identify the effects of air pollution. For example, Greenstone and Hanna (2014) show India's anti-pollution laws led to improvements in air pollution and had positive but insignificant impacts on infant mortality. Barrows et al. (2019) show the expansion of coal-fired power generation increases surrounding NO_2. One giga-watt increase in coal-fired capacity corresponds to a 14% rise in infant mortality rates in districts near versus far from the plant site. He et al. (2016) exploit exogenous variations in air quality during the 2008 Beijing Olympic Games, and find that a 10% decrease in PM_{10} concentrations reduces the monthly standardized all-cause mortality rate by 8%. Exploiting regression discontinuity designs based on the exact starting dates of winter heating across different cities, Fan et al. (2020) find turning on the winter heating system increased the weekly Air Quality Index by 36% in China and caused a 14% increase in mortality rate. Imelda (2018) study the effect of the Indonesian kerosene-to-LPG (liquid petroleum gas) conversion program. The program produces significantly less indoor air pollution and leads to 0.4 fewer infant deaths per thousand live births.

Another health effects lie in morbidity expenditure and hospital visits. Similar to mortality, most papers use similar identification strategies (weather conditions and environmental policies) to evaluate morbidity effects. For example, Deryugina et al. (2019) shows each $1\mu g/m^3$ increase in $PM_{2.5}$ increases three-day ER visits by 2.7 per million and inpatient ER spending by $16,000 per million population in the U.S. Also using weather variations for identification, Barwick et al. (2018) find a $10\mu g/m^3$ decrease in $PM_{2.5}$ reduces annual healthcare spending by more than $9.2 billion in China, 1.5% of the average annual healthcare expenditure. Deschenes et al. (2017)

estimate NO_x and O_3 dropped by 35% and 6% after the NO_x Budget Trading Program started. As a result, drug expenditures decreased by 1.9% in the participating states.

A small but growing number of papers provide novel instruments for air pollution and assess its health effects. For example, Schlenker and Walker (2015) use network delays to instrument air pollution in California airports. They find a one standard deviation increase in daily airplane taxi time increases CO by 23%, and a one standard deviation increase in pollution explains one third of daily asthma admissions. Relying on a driving restriction policy and cultural preference, Zhong et al. (2017) find 22% higher NO_2 due to traffic congestion increases ambulance calls by 12% in Beijing. Ito and Zhang (2020) find household's spending on air purifiers increases by \$1.34 as PM_{10} increases by $1\mu g/m^3$ induced by the Huai River heating policy.

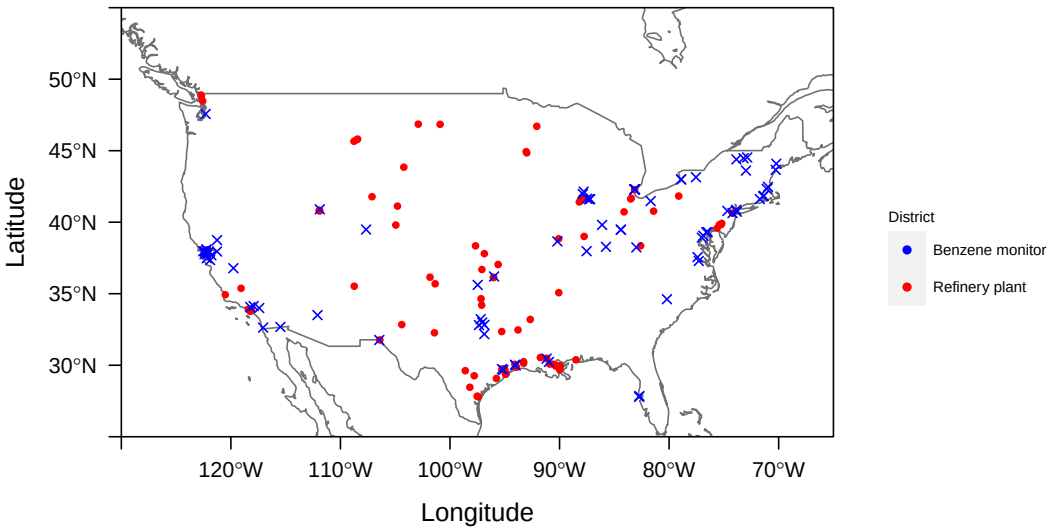
This paper adds to the pollution-health literature in three ways. First, I provide another innovative natural experiment of air pollution. My identification strategy does not rely on weather conditions that may naturally affect air quality and downstream outcomes. I also overcome the low power concern by exploring a common polluting problem. Second, given the short-lasting excess emissions and identified severe health effects, the paper highlights the importance of studying the short-term impacts of high polluting events which generally attract less discussion in the literature. Third, despite more medical expenditure and hospital visits, I find increasingly severe information asymmetry and insufficient avoidance behaviors in response to pollution events. This paper provides evidence that current environmental policies are ineffective in reducing pollution and alerting surrounding residents.

S3 Additional figures

Figure S3.1: Visual evidence of refinery pollution from the space

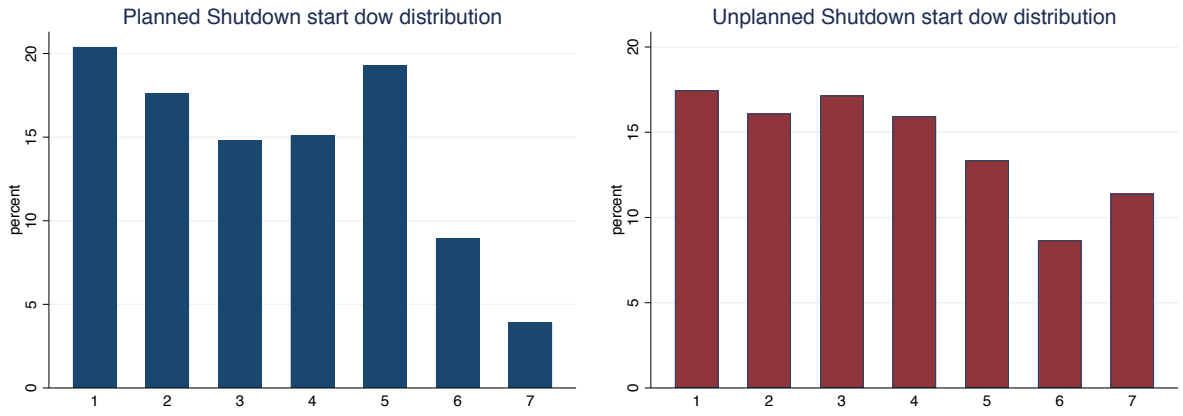
Notes: This figure shows satellite images of the Ferndale refinery plant on on November 2, 2019 (left) and March 29, 2019 (middle) when the plant was under normal operations. The right figure was taken on October 17, 2018 when the plant was under maintenance. Satellite images are from PlanetScope. Resolution is 3 meter. Cloud coverage is 0%.

Figure S3.2: Locations of refinery plants and benzene monitors

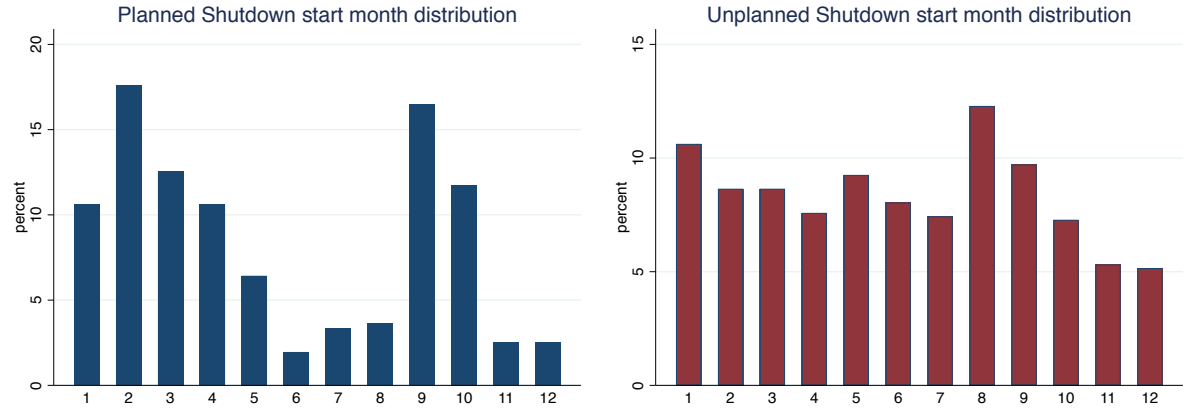


Notes: Locations of refinery plants (red) and benzene monitors (blue) are from the EIA and the EPA. While refineries are widely distributed across the country, monitors are concentrated in coastal and northeast areas

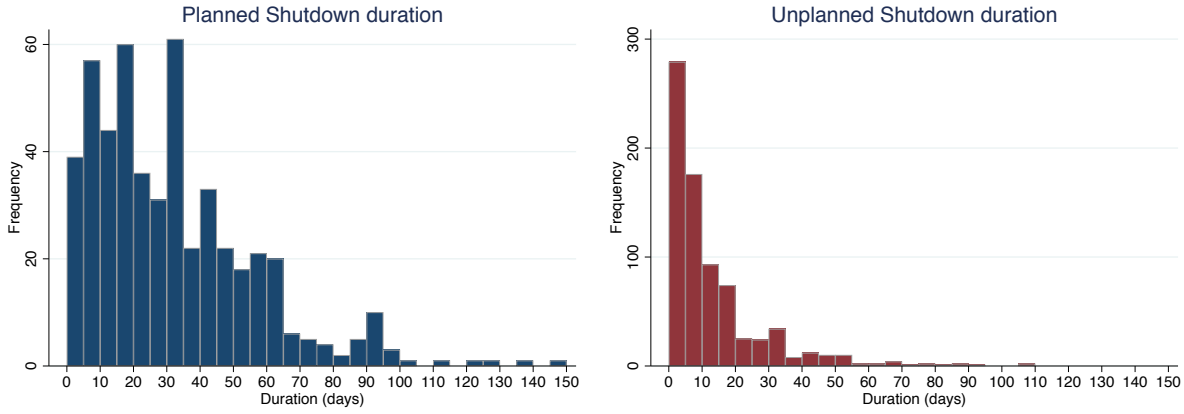
Figure S3.3: Summary figures of planned and unplanned outages' timings



(a) Distribution of planned and unplanned outage start time over the week

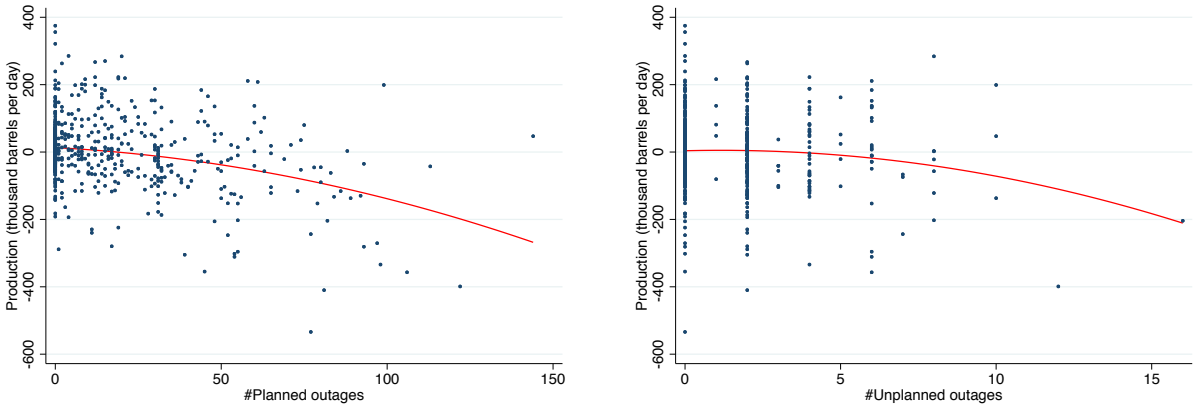


(b) Distribution of planned and unplanned outage start month



(c) Duration of planned and unplanned outages

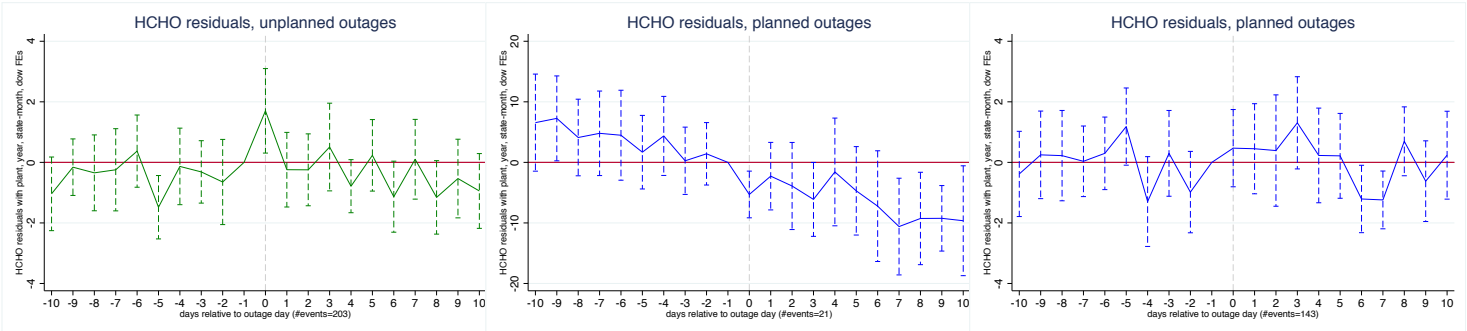
Figure S3.4: Outages and refinery production at the district-month level



(a) Planned outages (b) Unplanned outages

Notes: These figures display the raw outage count and production level in blue dots, and quadratic fitted curve in red.

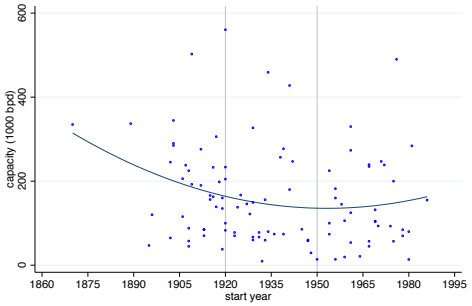
Figure S3.5: Effects of temporary planned and unplanned outages on HCHO



(a) Abnormal outages, all plants (b) Normal outages, big plants (c) Normal outages, all plants

Notes: These figures display the treatment effect of refinery plants' temporary unplanned and planned outages on surrounding HCHO levels in each day relative to the pre-treatment period (day -1). Panel (b) displays the effect of planned outages only using big plants with capacity above 200,000 barrels per day. In Panel (a) and (c), I use all plants without any capacity constraint, so the number of events reported in parentheses is larger than that in Panel (b) and Figure 2. All these three panels only use balanced events without any missing satellite reports in the 11-day event window. The green and blue solid lines display the estimated coefficients after controlling for plant, year, state-month and day of week fixed effects. The green and blue dash bars show 95% confidence intervals.

Figure S3.6: Plant age and capacity



Notes: This figure shows the nonlinear relationship between plant age and capacity. Blue scatter points display the raw capacity and start year of each plant, and the navy line shows the quadratic fitted relationship. Both old and new plants have large operation capacity (over 200K bpd), while plants constructed 1920-1950 (between the gray vertical lines) have smaller capacities.

Figure S3.7: Refinery plants with different crude oil importers

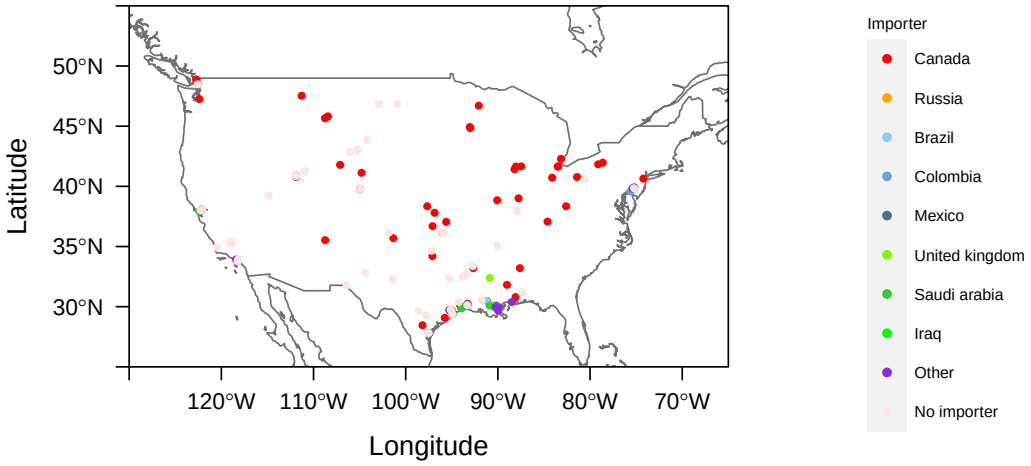
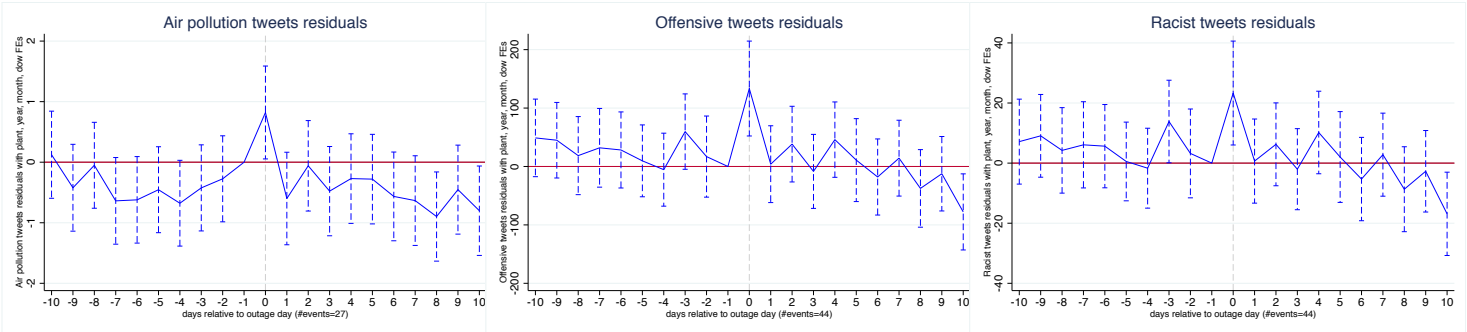


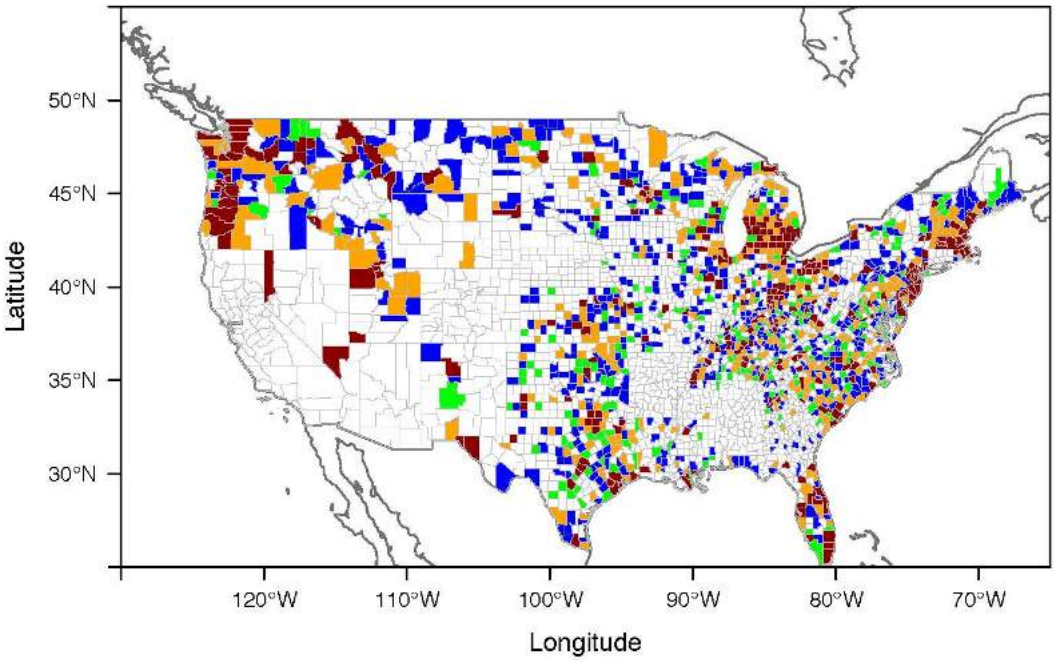
Figure S3.8: Effects of unplanned shutdowns on tweets with air pollution keywords, offensive tweets, and racist tweets



(a) Air pollution tweet (b) Offensive tweet (c) Racist tweet

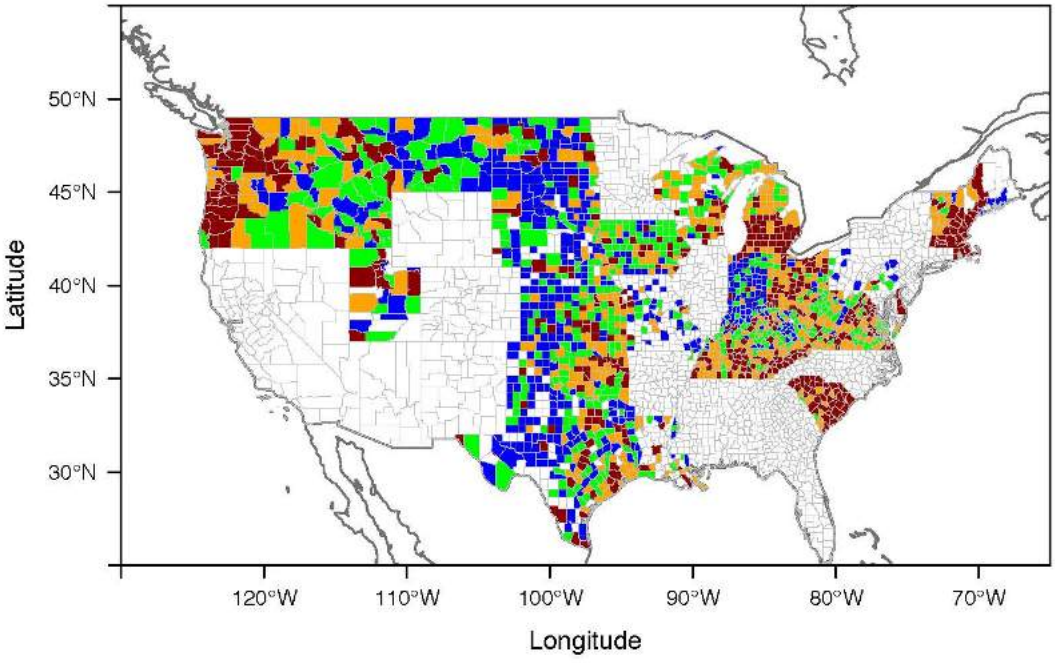
Notes: These figures display the treatment effect of refinery plants' temporary abnormal outages on tweet outcomes in each day relative to the pre-treatment period (day -1). The sample includes top ten refinery plants in Table S7.1. I only use balanced events without any missing satellite reports in the 11-day event window. The blue solid lines display the estimated coefficients after controlling for plant, year, month and day of week fixed effects. The blue dash bars show 95% confidence intervals.

Figure S3.9: Counties with hate crime reports 2014-2019



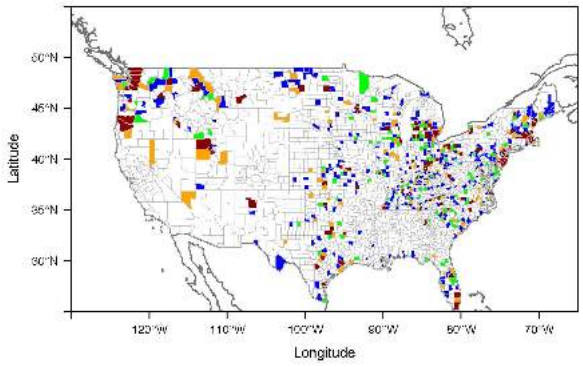
Notes: The number of hate crime events is shown in color. From bottom to top quantile, counties are filled with blue, green, orange, and dark red.

Figure S3.10: Counties with non-hate crime reports 2014-2019

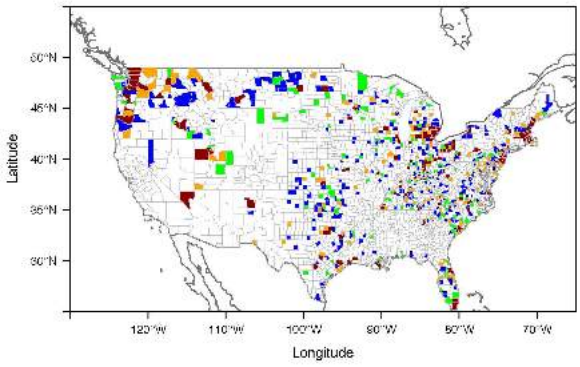


Notes: The number of non-hate crime events is shown in color. From bottom to top quantile, counties are filled with blue, green, orange, and dark red.

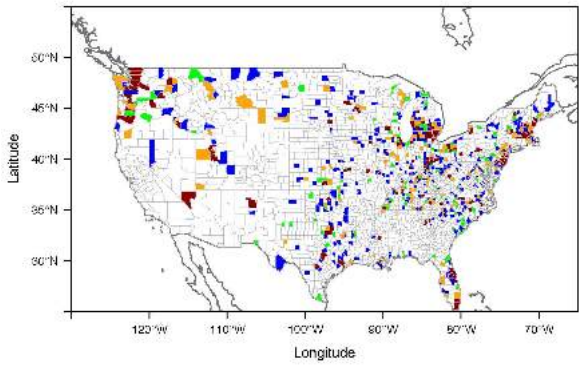
Figure S3.11: Counties with hate crime reports in each year



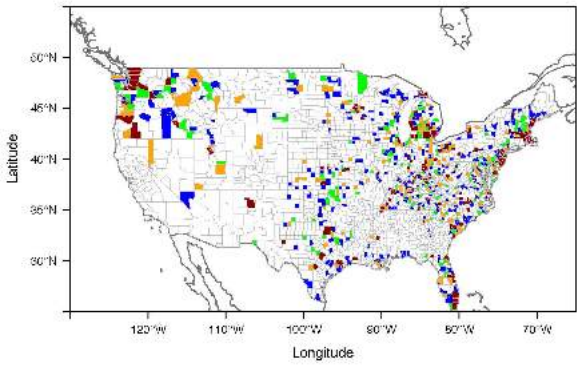
(a) 2014



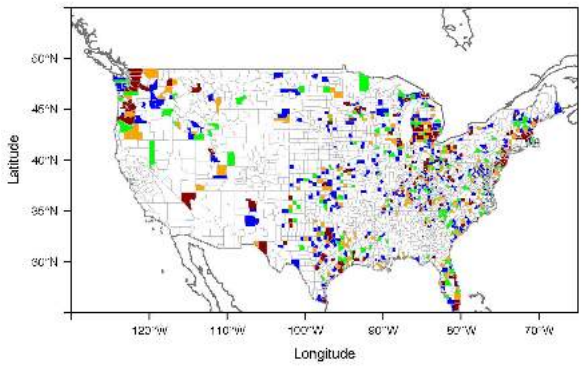
(b) 2015



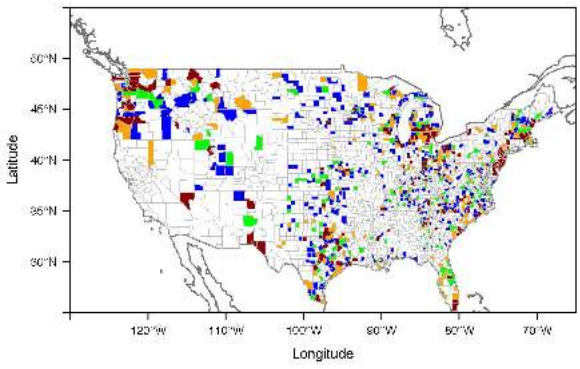
(c) 2016



(d) 2017

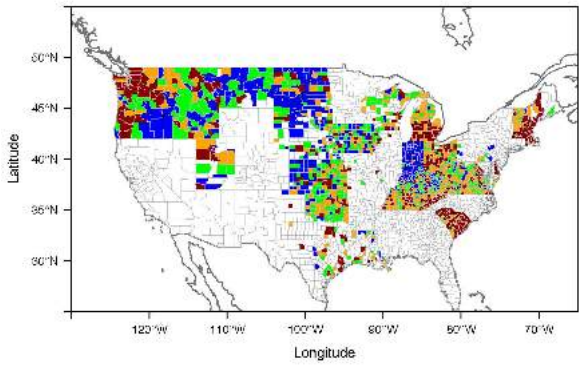


(e) 2018

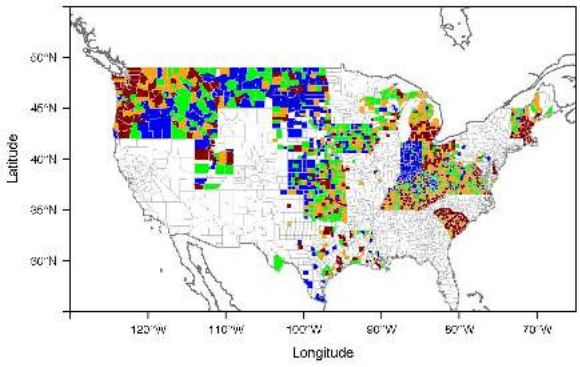


(f) 2019

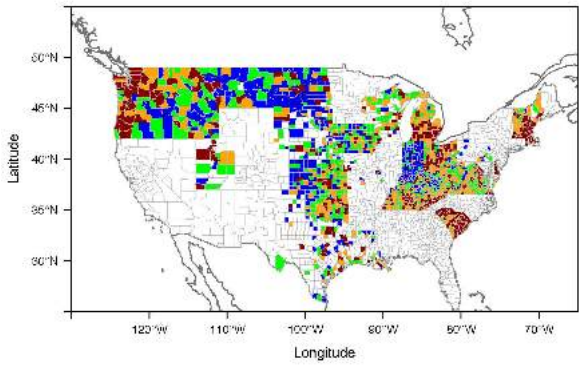
Figure S3.12: Counties with non-hate crime reports in each year



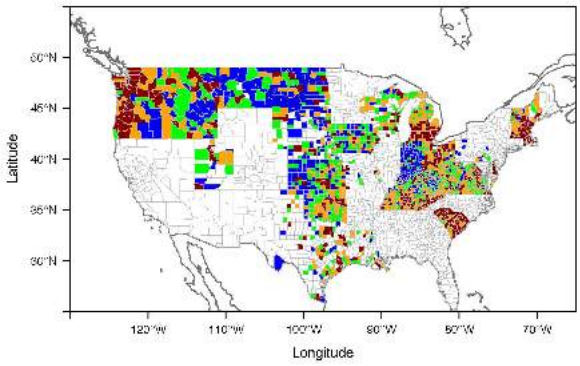
(a) 2014



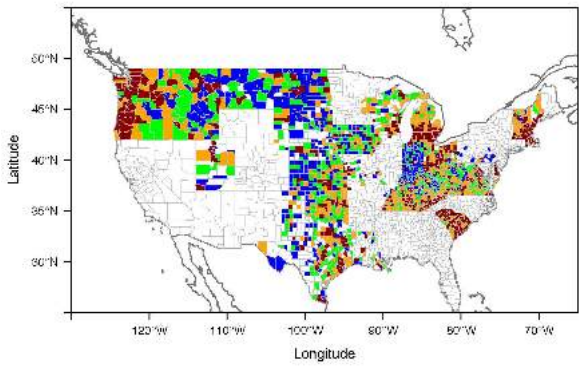
(b) 2015



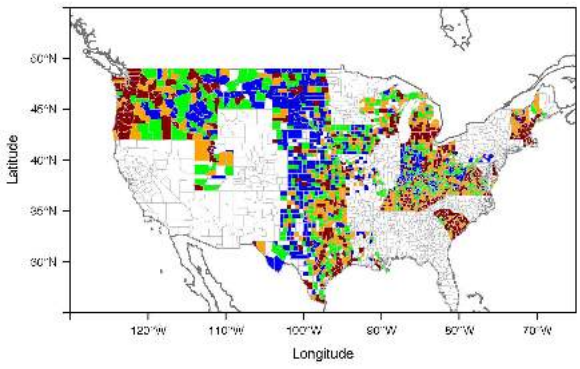
(c) 2016



(d) 2017



(e) 2018



(f) 2019

Figure S3.13: Word cloud of offensive tweets (left) and non-offensive tweets (right)

Notes: This figure plots the word cloud of nouns mentioned in the tweets classified as offensive and non-offensive close to refinery plants. The font size of each word is positively correlated with the frequency that the word is mentioned.

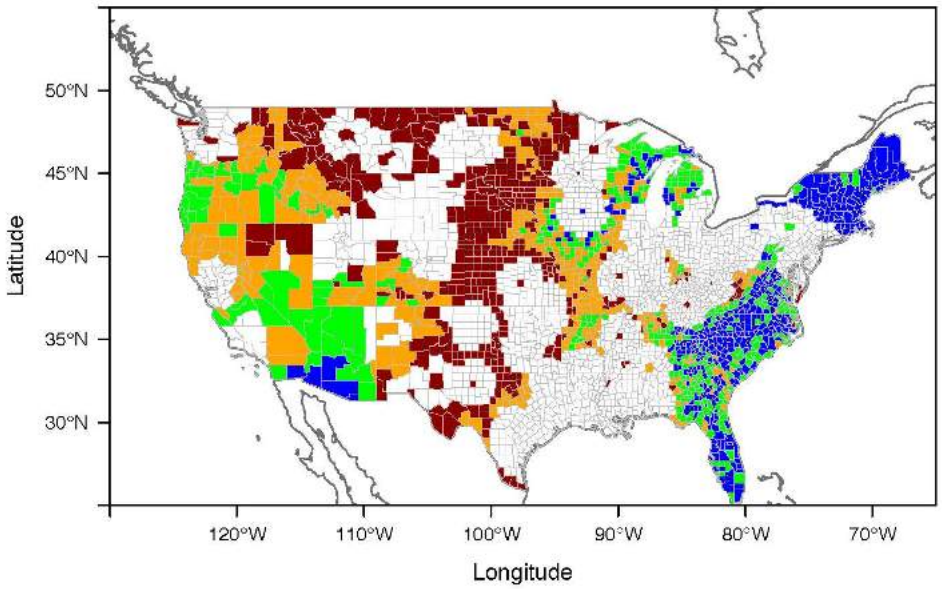
Figure S3.14: Word cloud of offensive tweets (left) and non-offensive tweets (right) after removing swear words

Notes: This figure plots the word cloud of nouns mentioned in the tweets classified as offensive and non-offensive close to refinery plants. The font size of each word is positively correlated with the frequency that the word is mentioned. Swear words are removed before constructing the frequency.

Figure S3.15: Word cloud of racist tweets (left) and non-racist tweets (right) after removing swear words

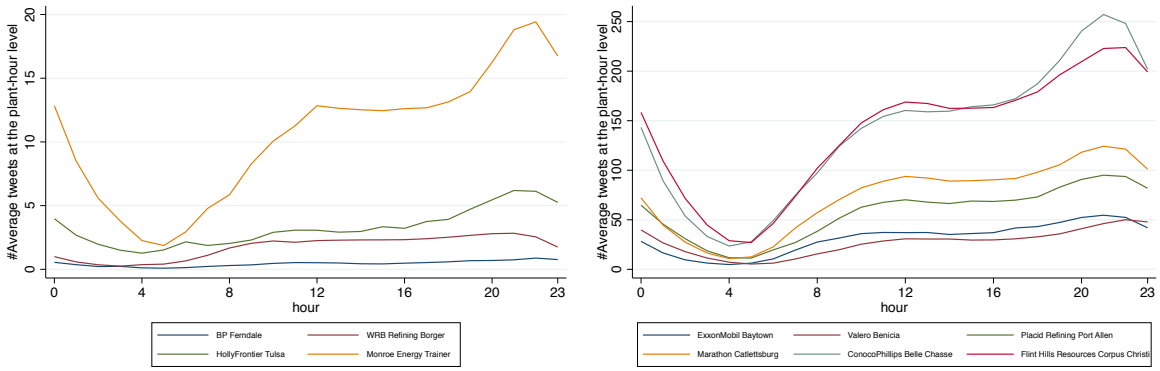
Notes: This figure plots the word cloud of nouns mentioned in the tweets classified as racist and non-racist close to refinery plants. The font size of each word is positively correlated with the frequency that the word is mentioned. Swear words are removed before constructing the frequency.

Figure S3.16: Geographically distant counties from refineries



Notes: This figure shows the spatial distribution of sample counties in Section 9. Red, orange, green, and blue counties have high, medium high, medium low, and low connectedness with refineries. Uncolored counties are within 150km of refineries and excluded in the analysis.

Figure S3.17: Diurnal cycles of tweeting activities near top ten refinery plants



S4 Additional tables on the first stage

Table S4.1: Correlation of satellite HCHO and ground-based benzene

	HCHO	
	(1)	(2)
Benzene	0.151***	0.186***
	(0.035)	(0.043)
Observations	12562	12562
R-square	0.158	0.546
Y-mean	9.014	9.014
Y-sd	6.025	6.025
Monitor FEs	Y	Y
Year FEs	Y	
Month FEs	Y	
DOW FEs	Y	
Day FEs		Y

Notes: The sample includes HCHO near monitors and benzene readings from monitors. I use the average HCHO near 20km buffer of each monitor.

Table S4.2: Duration of planned and unplanned outages

	Planned outages	Unplanned outages
Within 1 week	67	384
	(0.13)	(0.5)
1-2 weeks	73	164
	(0.14)	(0.21)
2-3 weeks	77	81
	(0.15)	(0.1)
3-4 weeks	37	34
	(0.07)	(0.04)
4-5 weeks	76	45
	(0.15)	(0.05)
Over 5 weeks	175	54
	(0.34)	(0.07)
Total events	505	762
Mean duration (days)	31.26	12.50
Std. dev.	24.17	14.97

Notes: The fraction of events in each duration bin is reported in parentheses.

Table S4.3: Outages and refinery production

	Panel A: Production (thousand barrels per day)		
	(1)	(2)	(3)
#Unplanned outages	-3.164*	-3.546**	-3.703**
	(1.772)	(1.715)	(1.787)
#Planned outages	-2.861***	-2.645***	-2.707***
	(0.176)	(0.179)	(0.184)
Observations	864	864	864
R-square	0.993	0.994	0.994
Y-mean	988.631	988.631	988.631
Y-sd	1009.888	1009.888	1009.888
	Panel B: Only unplanned outages and production		
#Unplanned outages	-11.422***	-10.918***	-11.394***
	(1.942)	(1.840)	(1.932)
Observations	864	864	864
R-square	0.991	0.992	0.992
Y-mean	988.631	988.631	988.631
Y-sd	1009.888	1009.888	1009.888
	Panel C: Only planned outages and production		
#Planned outages	-2.951***	-2.753***	-2.818***
	(0.169)	(0.172)	(0.176)
Observations	864	864	864
R-square	0.993	0.994	0.994
Y-mean	988.631	988.631	988.631
Y-sd	1009.888	1009.888	1009.888
District FEs	Y	Y	Y
Year FEs		Y	
Month FEs		Y	
Year-month FEs			Y

Notes: The correlation test is at the district-month level. There are 12 districts in the U.S. and 72 months in the sample period. Standard errors are clustered at the district level.

Table S4.4: Effects of temporary unplanned and planned outages on HCHO

	HCHO			
	Panel A: Unplanned outages			
UnplannedShut	0.584**	0.594**	0.583**	0.584**
	(0.243)	(0.248)	(0.242)	(0.242)
UnplannedDowntime	-0.221*	-0.208*	-0.223*	-0.224*
	(0.110)	(0.109)	(0.110)	(0.111)
OutageAnyPlant	0.028	-0.002	0.010	-0.002
	(0.105)	(0.171)	(0.104)	(0.108)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
	Panel B: Planned outages			
	(1)	(2)	(3)	(4)
PlannedShut	0.028	-0.048	0.025	0.025
	(0.223)	(0.218)	(0.224)	(0.224)
PlannedDowntime	-0.457***	-0.500***	-0.458***	-0.459***
	(0.080)	(0.076)	(0.080)	(0.079)
OutageAnyPlant	0.029	0.000	0.011	-0.002
	(0.105)	(0.171)	(0.105)	(0.108)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

S5 Heterogeneity across outage time and plant characersitics

In this section, I discuss the heterogeneous effects of abnormal outages on pollution across outage timing and plant characteristics. First, motivated by the distribution of the start day of week in Figure S3.3, I check whether abnormal operations on weekends generate more severe pollutant releases. To do so, I interact $UnplannedShut$ with weekday dummy and weekend dummy in equation (1). Since day of week fixed effects are already controlled, this practice serves as a difference-in-difference test to estimate the extra impact of weekend unplanned outages. Results in Table S5.1 show positive and significant estimates on $UnplannedShut \times Weekday$. The magnitude is similar to those in Table 2. The abnormal shutdown increases surrounding HCHO levels by 0.54 units, 6.6% relative to the mean. Estimates on $UnplannedShut \times Weekend$ are not statistically significant but 114% larger than estimates on $UnplannedShut \times Weekday$. The imprecise estimates may be due to the smaller sample size of the weekend treatment group. Given the large point estimates, unplanned outages starting on weekends have larger pollution increases than those starting on weekdays. The effect is 1.2 units HCHO increase, 14.2% relative to the mean. The release of pollutants or catalytic may be controlled to some degree on weekdays when more workers are on guard. As a result, unplanned outages are more polluting and harmful on weekends, so plant operators tend to shut down plants on weekdays though unexpected equipment failures are quasi-random in time.

Second, I test the heterogeneity across plant capacities. Larger plants process more crude oil and generate more significant environmental impacts than smaller plants. Given the environmental concern, no new complex refineries over 100,000 bpd have been built since the 1970s (Chesnes, 2015). If the pollution impact of normal operation is more remarkable for larger plants, I hypothesize the excess pollutant emission when unplanned outages are also more severe. Therefore, I separate plants into three groups, below 100,000 bpd, above 200,000 bpd, and those in between. I re-estimate equation (1) for each group. Results in Table S5 show increasing point estimates on $UnplannedShut$ from Panel A to Panel C. Plants with medium and smaller capacity have similar environmental impacts, while those with larger capacity above 200,000 bpd release more pollutants when unplanned shutdowns. HCHO increases for each group are estimated as 0.6, 0.6 and 1.3 units, 7.7%, 7.7% and 15.9% relative to the mean. The capacity-pollution relationship is not linear but shows a discontinuous increase at a capacity threshold. For plants below 200,000 bpd, environmental impacts of abnormal operations are similar. As capacity exceeds 200,000 bpd, the effects become much more potent. Besides, the average HCHO values are higher near larger plants than smaller plants, 8.5 vs. 7.9 units. This indicates the generally worse environmental conditions near refineries with capacity above 200,000 bpd.

Third, I check whether older plants are less controlled for unexpected pollutant release. I separate plants into three groups, those built before 1920, 1920-1950, and after 1950. Table S5.3 shows positive estimates on $UnplannedShut$ in Panel A and C, while estimates are small and

negative in Panel B. Unlike capacity, age and pollution have a non-monotonic relationship. Both old and new plants' abnormal outages result in severe pollution increases. Specifically, plants older than 100 years old lead to 0.9 units (10.4%) increase in nearby HCHO on the unplanned shutdown day. Those younger than 70 years old lead to 1.2 units (15.3%) HCHO increase. The finding in Panel A suggests old plants tend to have poorly designed control rooms. The average VOC level is the highest in Panel A, indicating the worst air quality near the oldest group of refinery plants. For new plants, they may share a higher burden of oil product supply and operate at a higher utilization rate. Besides, plant age and capacity are highly correlated, as is shown in Figure S3.6. There is a nonlinear relationship between plant age and capacity. Older and newer plants tend to have larger capacities than plants built 1920-1950. It could be either age or capacity that leads to the heterogeneous effect of abnormal operations.

Besides, I assess heterogeneous effects across crude oil sources. I hypothesize that surrounding air quality is worse if crude oil is of poorer quality. Data on plant-level imports is from EIA-814, Monthly Imports Report. 79 out of my 101 sample refinery plants have importer information, including processing company name, processing facility name, import country, import quantity, and sulfur percent. For the other 22 plants without importer records, I assume they only use domestically produced crude oil as inputs. Figure S3.7 summarizes the dominant importer of each plant. 51 (35.2%) plants' dominant importers are Canada. Other big importing countries include Saudi Arabia (8 plants), Iraq (6), Mexico (5), Russia (3), and Colombia (3). I estimate equation (1) using subgroups of refineries with domestically produced and imported oil and separate by specific importers. Results in Table S5.4 show effects of unplanned outages are driven by plants using imported crude oil. Plants refining domestic oil have no significant impacts on surrounding air quality when abnormal shutdowns, but still generate pollution when normal operations, captured by negative estimates on $Downtime$. Adding importer country fixed effects has no impact on estimates of interest or model performances, shown in similar R^2s. In Table S5.5, among foreign importers, refineries using crude oil from Brazil, U.K., and Iraq see higher air pollution spikes when abnormal outages, while those importing from Canada and Mexico experience lower pollution increases. I further explore the heterogeneity across imported oil's sulfur content. In Table S5.6 Column (2), low sulfur contents are correlated with high pollution increases, captured by the negative estimate on $UnplannedShut \times Sulfur$. Similar patterns are found in Column (3)-(5). Despite lower pollution spikes, refineries using higher sulfur crude oil have worse air quality in surrounding areas in general, as shown in Y-mean. I conclude that longer shipment distances are associated with higher pollution spikes, while the severity of pollutant releases in abnormal operations does not result from poor fuel quality.

Table S5.1: Heterogeneous effects on weekdays and weekends

	HCHO			
	(1)	(2)	(3)	(4)
UnplannedShut × Weekday	0.543**	0.548**	0.540**	0.540**
	(0.248)	(0.252)	(0.248)	(0.248)
UnplannedShut × Weekend	1.134	1.184	1.153	1.157
	(0.751)	(0.745)	(0.756)	(0.757)
Downtime × Weekday	-0.333***	-0.358***	-0.335***	-0.336***
	(0.053)	(0.050)	(0.054)	(0.053)
Downtime × Weekend	-0.419**	-0.360*	-0.423**	-0.422**
	(0.198)	(0.193)	(0.199)	(0.199)
OutageAnyPlant	.0504	.0211	.0353	.0231
	(.0974)	(.154)	(.0977)	(.101)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S5.2: Heterogeneous effects across plant capacity

	HCHO			
	Panel A: Plants with capacity <100,000 bpd			
	(1)	(2)	(3)	(4)
UnplannedShut	0.608**	0.576**	0.606**	0.609**
	(0.271)	(0.269)	(0.264)	(0.266)
Downtime	-0.382	-0.351	-0.382	-0.384
	(0.398)	(0.381)	(0.397)	(0.396)
OutageAnyPlant	0.075	0.072	0.075	0.064
	(0.044)	(0.111)	(0.044)	(0.043)
Observations	42940	42940	42940	42940
R-square	0.118	0.121	0.118	0.118
Y-mean	7.875	7.875	7.875	7.875
Y-sd	5.496	5.496	5.496	5.496
	Panel B: Plants with capacity 100,000 to 200,000 bpd			
UnplannedShut	0.610**	0.635**	0.616**	0.613**
	(0.262)	(0.252)	(0.265)	(0.265)
Downtime	-0.374***	-0.389***	-0.375***	-0.373***
	(0.091)	(0.073)	(0.092)	(0.091)

OutageAnyPlant	0.033	0.002	0.030	0.022
	(0.089)	(0.111)	(0.090)	(0.093)
Observations	30632	30632	30632	30632
R-square	0.134	0.139	0.135	0.135
Y-mean	7.890	7.890	7.890	7.890
Y-sd	5.399	5.399	5.399	5.399
	Panel C: Plants with capacity ≥200,000 bpd			
UnplannedShut	1.347**	1.368**	1.345**	1.353**
	(0.448)	(0.406)	(0.447)	(0.449)
Downtime	-0.210**	-0.238*	-0.211*	-0.211*
	(0.088)	(0.114)	(0.089)	(0.094)
OutageAnyPlant	-0.123	-0.134	-0.125	-0.156
	(0.172)	(0.270)	(0.170)	(0.181)
Observations	21206	21206	21206	21206
R-square	0.124	0.131	0.124	0.124
Y-mean	8.467	8.467	8.467	8.467
Y-sd	5.694	5.694	5.694	5.694
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S5.3: Heterogeneous effects across plant age

	HCHO			
	Panel A: Plants starting before 1920			
	(1)	(2)	(3)	(4)
UnplannedShut	0.885**	0.910**	0.889**	0.889**
	(0.370)	(0.364)	(0.369)	(0.370)
Downtime	-0.346**	-0.378***	-0.349**	-0.348**
	(0.133)	(0.120)	(0.132)	(0.133)
OutageAnyPlant	-0.213	-0.570*	-0.225*	-0.223*
	(0.119)	(0.300)	(0.116)	(0.116)
Observations	27729	27729	27729	27729
R-square	0.132	0.138	0.132	0.132
Y-mean	8.521	8.521	8.521	8.521
Y-sd	5.791	5.791	5.791	5.791
	Panel B: Plants starting 1920-1950			
UnplannedShut	-0.231	-0.212	-0.232	-0.230
	(0.715)	(0.733)	(0.712)	(0.713)
Downtime	-0.149	-0.188*	-0.146	-0.152
	(0.099)	(0.101)	(0.099)	(0.103)
OutageAnyPlant	-0.092	-0.082	-0.110	-0.117
	(0.092)	(0.098)	(0.087)	(0.093)

Observations	41717	41717	41717	41717
R-square	0.116	0.120	0.116	0.116
Y-mean	8.128	8.128	8.128	8.128
Y-sd	5.534	5.534	5.534	5.534
	Panel C: Plants starting after 1950			
UnplannedShut	1.201**	1.205**	1.201**	1.201**
	(0.534)	(0.550)	(0.539)	(0.539)
Downtime	-0.522*	-0.544*	-0.524*	-0.524*
	(0.260)	(0.277)	(0.261)	(0.262)
OutageAnyPlant	-0.105	-0.120	-0.110	-0.112
	(0.085)	(0.080)	(0.084)	(0.082)
Observations	35889	35889	35889	35889
R-square	0.138	0.141	0.138	0.138
Y-mean	7.829	7.829	7.829	7.829
Y-sd	5.448	5.448	5.448	5.448
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S5.4: Heterogeneous effects by whether or not crude oil is imported

	HCHO		Refineries using		
			domestic crude oil	imported crude oil	
UnplannedShut	0.605**	0.605**	0.551	0.662**	0.662**
	(0.246)	(0.246)	(0.496)	(0.288)	(0.288)
Downtime	-0.355***	-0.355***	-0.351***	-0.327***	-0.327***
	(0.054)	(0.054)	(0.091)	(0.072)	(0.072)
Observations	105335	105335	24963	80372	80372
R-square	0.128	0.128	0.134	0.123	0.123
Y-mean	8.129	8.129	7.703	8.368	8.368
Y-sd	5.581	5.581	5.242	5.735	5.735
Plant FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month DOW FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Domestic importer FEs		Y			Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S5.5: Heterogeneous effects across crude oil importers

	HCHO			
	Canada	Russia	Brazil	Colombia
	(1)	(2)	(3)	(4)
UnplannedShut	0.766**	0.857**	1.236***	0.931**
	(0.289)	(0.279)	(0.209)	(0.255)
Downtime	-0.323***	-0.364***	-0.228**	-0.385***
	(0.057)	(0.071)	(0.066)	(0.087)
Observations	74597	39691	38779	36495
R-square	0.119	0.133	0.139	0.131
Y-mean	8.358	8.438	8.558	8.420
Y-sd	5.717	5.675	5.743	5.618
	Mexico	UK	Saudi Arabia	Iraq
UnplannedShut	0.877*	1.024***	0.912*	1.019**
	(0.442)	(0.204)	(0.456)	(0.366)
Downtime	-0.348***	-0.215**	-0.297***	-0.272***
	(0.068)	(0.066)	(0.053)	(0.056)
Observations	34646	34156	34093	33725
R-square	0.131	0.144	0.137	0.128
Y-mean	8.493	8.793	8.524	8.517
Y-sd	5.667	5.907	5.738	5.691
Plant FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month DOW FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S5.6: Heterogeneous effects across crude oil's sulfur content

	HCHO		Subsample, crude oil's sulfur percent		
			$(0, 1]$	$(1, 2]$	$(2, 3.1]$
UnplannedShut	0.783**	2.375***	1.380**	1.099**	-0.163
	(0.284)	(0.690)	(0.529)	(0.442)	(0.456)
UnplannedShut × Sulfur		-0.778**			
		(0.366)			
Downtime	-0.322***	-0.323***	-0.635***	-0.199***	-0.211**
	(0.054)	(0.053)	(0.200)	(0.044)	(0.092)
Observations	80372	80372	27469	26792	26111
R-square	0.124	0.124	0.153	0.108	0.110
Y-mean	8.336	8.336	8.390	8.083	8.539
Y-sd	5.703	5.703	5.883	5.421	5.784
Plant FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month DOW FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

S6 Robustness

Table S6.1: Adding weather controls

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	0.068	0.010	0.063	0.063
	(0.244)	(0.230)	(0.243)	(0.243)
PlannedDowntime	-0.369***	-0.409***	-0.370***	-0.372***
	(0.078)	(0.078)	(0.079)	(0.078)
UnplannedShut	0.542*	0.547*	0.539*	0.539*
	(0.285)	(0.289)	(0.284)	(0.284)
UnplannedDowntime	-0.306***	-0.299***	-0.308***	-0.311***
	(0.062)	(0.068)	(0.063)	(0.063)
OutageAnyPlant	-.0708	-.0867	-.0874	-.104
	(.114)	(.141)	(.117)	(.109)
Temperature	-.00268	-.00579	-.00227	-.00257
	(.022)	(.0213)	(.0221)	(.0222)
WindSpeed	.0223*	.0238*	.0224*	.0228*
	(.0108)	(.0124)	(.0109)	(.0112)
Precipitation	.00146	.00145	.00138	.0014
	(.00585)	(.00588)	(.00584)	(.00584)
Observations	92138	92138	92138	92138
R-square	0.120	0.124	0.121	0.121
Y-mean	8.154	8.154	8.154	8.154
Y-sd	5.574	5.574	5.574	5.574
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.2: Predictors of unplanned shutdowns

	Occurrence of unplanned shutdown			
	OLS			
	(1)	(2)	(3)	(4)
Temperature	0.129***	0.021	0.024	0.022
	(0.039)	(0.029)	(0.027)	(0.030)
WindSpeed	-0.037	-0.030	-0.034	-0.030
	(0.058)	(0.054)	(0.057)	(0.050)
Precipitation	0.000	-0.015	-0.013	-0.015
	(0.014)	(0.016)	(0.015)	(0.016)
Observations	92138	92138	92138	92138
R-square	0.001	0.006	0.007	0.006
Y-mean	2.898	2.898	2.898	2.898
Y-sd	53.754	53.754	53.754	53.754

	Logit			
Temperature	0.006	0.006	0.007	0.006
	(0.010)	(0.010)	(0.010)	(0.010)
WindSpeed	-0.011	-0.011	-0.013	-0.011
	(0.018)	(0.018)	(0.019)	(0.017)
Precipitation	-0.004	-0.004	-0.002	-0.004
	(0.006)	(0.006)	(0.005)	(0.006)
Observations	92138	92138	92138	92138
Pseudo R-square	0.067	0.067	0.081	0.067
Y-mean	4.677	4.677	4.862	4.677
Y-sd	68.228	68.228	69.561	68.228
Plant FEs		Y	Y	Y
Year FEs	Y	Y		Y
Month FEs	Y	Y		Y
Year-month FEs			Y	
DOW FEs	Y	Y	Y	Y
Trends				Quadratic

Notes: Occurrence dummy is multiplied by 1000. Standard errors are clustered at the plant level.

Table S6.3: Dropping weather forecast-induced outages

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	-0.014	-0.090	-0.017	-0.016
	(0.226)	(0.220)	(0.226)	(0.227)
PlannedDowntime	-0.449***	-0.493***	-0.450***	-0.451***
	(0.079)	(0.075)	(0.080)	(0.079)
UnplannedShut	0.597**	0.611**	0.597**	0.597**
	(0.250)	(0.257)	(0.250)	(0.250)
UnplannedDowntime	-0.164	-0.146	-0.165	-0.166
	(0.118)	(0.116)	(0.118)	(0.118)
OutageAnyPlant	-.0766	-.103	-.0936	-.106
	(.104)	(.13)	(.105)	(.102)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

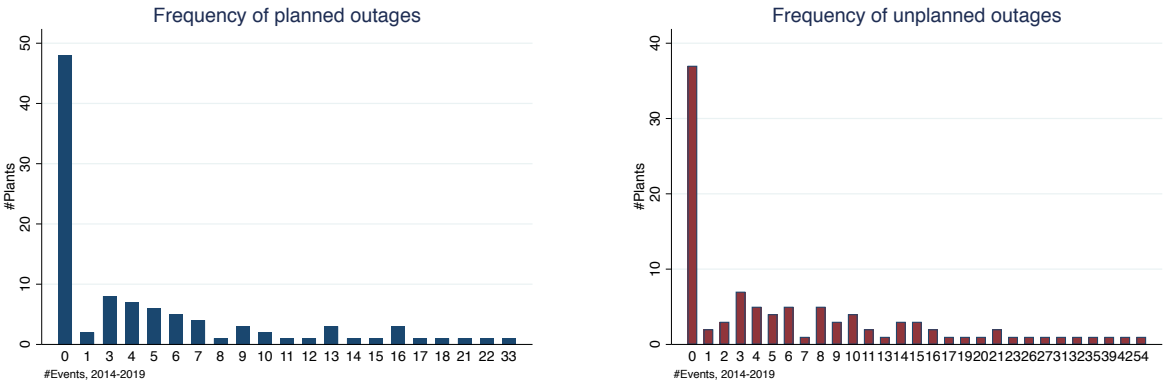
Notes: Standard errors are clustered at the plant level.

Table S6.4: Dropping outages with capacity offline ≤50%

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	0.312	0.266	0.299	0.295
	(0.524)	(0.515)	(0.517)	(0.516)
PlannedDowntime	-0.376***	-0.454***	-0.378***	-0.383***
	(0.097)	(0.102)	(0.096)	(0.096)
UnplannedShut	1.161*	1.174*	1.159*	1.160*
	(0.669)	(0.680)	(0.674)	(0.675)
UnplannedDowntime	-0.226	-0.192	-0.229	-0.229
	(0.158)	(0.149)	(0.157)	(0.157)
OutageAnyPlant	-.19**	-.292**	-.195**	-.212**
	(.0772)	(.136)	(.0778)	(.0797)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Figure S6.1: Frequency of planned and unplanned outages at each plant, 2014-2019



(a) Planned outages

(b) Unplanned outages

Table S6.5: Dropping plants without unplanned outages

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	0.027	-0.065	0.024	0.024
	(0.230)	(0.221)	(0.229)	(0.230)
PlannedDowntime	-0.388***	-0.439***	-0.390***	-0.390***
	(0.081)	(0.073)	(0.082)	(0.081)
UnplannedShut	0.601**	0.619**	0.600**	0.600**
	(0.247)	(0.252)	(0.247)	(0.247)
UnplannedDowntime	-0.161	-0.134	-0.163	-0.163
	(0.103)	(0.104)	(0.103)	(0.102)
OutageAnyPlant	-.0924	-.185*	-.109	-.114
	(.1)	(.0988)	(.107)	(.098)
Observations	75920	75920	75920	75920
R-square	0.131	0.135	0.132	0.132
Y-mean	8.404	8.404	8.404	8.404
Y-sd	5.667	5.667	5.667	5.667
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.6: Dropping plants with multiple abnormal outages

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	0.095	-0.101	0.104	0.113
	(1.027)	(1.055)	(1.018)	(1.015)
PlannedDowntime	-0.387	-0.497**	-0.385	-0.393
	(0.296)	(0.234)	(0.294)	(0.296)
UnplannedShut	2.078*	1.887*	2.097*	2.105*
	(1.162)	(1.088)	(1.175)	(1.176)
UnplannedDowntime	-0.979*	-0.933*	-0.961*	-0.949*
	(0.483)	(0.485)	(0.477)	(0.480)
OutageAnyPlant	.00991	.0476	-.00816	-.0286
	(.149)	(.178)	(.148)	(.14)
Observations	40977	40977	40977	40977
R-square	0.098	0.101	0.098	0.098
Y-mean	7.834	7.834	7.834	7.834
Y-sd	5.498	5.498	5.498	5.498
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.7: Dropping days after the second abnormal outage

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	-0.197	-0.264	-0.198	-0.195
	(0.336)	(0.336)	(0.335)	(0.337)
PlannedDowntime	-0.408***	-0.455***	-0.409***	-0.410***
	(0.108)	(0.095)	(0.108)	(0.107)
UnplannedShut	1.231**	1.160**	1.231**	1.238**
	(0.463)	(0.465)	(0.472)	(0.472)
UnplannedDowntime	-0.641***	-0.605***	-0.643***	-0.637***
	(0.154)	(0.141)	(0.155)	(0.156)
OutageAnyPlant	.128	.166	.112	.0893
	(.162)	(.146)	(.164)	(.154)
Observations	72789	72789	72789	72789
R-square	0.125	0.128	0.125	0.125
Y-mean	7.894	7.894	7.894	7.894
Y-sd	5.465	5.465	5.465	5.465
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.8: Decomposition of event study estimates

	Weight	β
Earlier T vs. Later C	0.255	0.536
Later T vs. Earlier C	0.001	0.542
T vs. Never treated	0.744	0.546
DD coefficient		0.543

Notes: There are some random missing reports in satellite VOC products. As `bacondecomp` requires a strongly balanced sample, I fill in missing values after estimating equation (1) and perform decomposition using the partially predicted sample. Estimates are slightly smaller than those in Table 2.

Figure S6.2: Decomposition of event study estimates

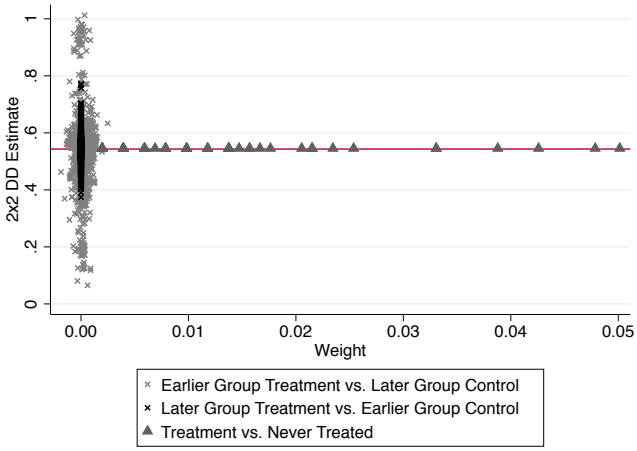


Table S6.9: Adding restarting effects

	HCHO			
	(1)	(2)	(3)	(4)
Restart_Planned	0.662	0.678	0.657	0.659
	(0.417)	(0.439)	(0.419)	(0.420)
Restart_Unplanned	0.146	0.161	0.149	0.150
	(0.430)	(0.423)	(0.431)	(0.430)
PlannedShut	-0.017	-0.094	-0.021	-0.020
	(0.225)	(0.219)	(0.225)	(0.225)
PlannedDowntime	-0.451***	-0.495***	-0.452***	-0.453***
	(0.078)	(0.074)	(0.079)	(0.078)
UnplannedShut	.597**	.61**	.596**	.597**
	(.249)	(.254)	(.249)	(.249)
UnplannedDowntime	-.192*	-.176	-.194*	-.195*
	(.107)	(.106)	(.107)	(.107)
OutageAnyPlant	-.0757	-.103	-.0926	-.105
	(.103)	(.13)	(.105)	(.102)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.10: Only isolated plants

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	0.013	-0.072	0.009	0.009
	(0.232)	(0.221)	(0.232)	(0.232)
PlannedDowntime	-0.365***	-0.410***	-0.366***	-0.366***
	(0.089)	(0.087)	(0.090)	(0.089)
UnplannedShut	0.684**	0.697**	0.683**	0.683**
	(0.272)	(0.278)	(0.271)	(0.271)
UnplannedDowntime	-0.179	-0.155	-0.181*	-0.182*
	(0.104)	(0.104)	(0.103)	(0.103)
OutageAnyPlant	.0746	-.0304	.0599	.0564
	(.0715)	(.091)	(.0776)	(.0666)
Observations	73760	73760	73760	73760
R-square	0.135	0.139	0.135	0.135
Y-mean	8.427	8.427	8.427	8.427
Y-sd	5.685	5.685	5.685	5.685
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.11: Effects of unplanned shutdowns on benzene from ground monitors

	Benzene			
	(1)	(2)	(3)	(4)
UnplannedShut	0.114**	0.137***	0.115**	0.114**
	(0.046)	(0.034)	(0.046)	(0.047)
Downtime	-0.049**	-0.060	-0.048***	-0.048***
	(0.016)	(0.035)	(0.016)	(0.015)
OutageAnyPlant	0.016	0.037*	0.017	0.018
	(0.017)	(0.018)	(0.016)	(0.015)
Observations	18807	18807	18807	18807
R-square	0.188	0.236	0.188	0.188
Y-mean	1.451	1.451	1.451	1.451
Y-sd	1.599	1.599	1.599	1.599
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.12: Effects of unplanned shutdowns on satellite AOD

	AOD			
	(1)	(2)	(3)	(4)
UnplannedShut	0.018**	0.017**	0.018**	0.017**
	(0.007)	(0.007)	(0.007)	(0.007)
Downtime	-0.003	-0.003	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)
OutageAnyPlant	0.012***	0.002	0.011***	0.013***
	(0.003)	(0.003)	(0.003)	(0.002)
Observations	79113	79113	79113	79113
R-square	0.227	0.259	0.227	0.228
Y-mean	0.148	0.148	0.148	0.148
Y-sd	0.139	0.139	0.139	0.139
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.13: Effects of unplanned shutdowns on other air pollutants

	SO_2	NO_2	CO
	(1)	(2)	(3)
UnplannedShut	0.063	0.819***	0.019*
	(0.039)	(0.152)	(0.010)
Observations	81637	51987	48999
R-square	0.243	0.625	0.314
Y-mean	0.963	8.460	0.278
Y-sd	1.618	4.877	0.147
Plant FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S6.14: Using HCHO at the county-day level

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	-0.334	-0.389	-0.332	-0.332
	(0.468)	(0.458)	(0.469)	(0.469)
PlannedDowntime	-0.331***	-0.356***	-0.327***	-0.327***
	(0.099)	(0.107)	(0.098)	(0.098)
UnplannedShut	0.825***	0.847***	0.825***	0.824***
	(0.279)	(0.288)	(0.280)	(0.280)
UnplannedDowntime	-0.155	-0.165	-0.153	-0.152
	(0.142)	(0.145)	(0.140)	(0.140)
OutageAnyPlant	-.0421	-.0837	-.0555	-.0496
	(.126)	(.115)	(.128)	(.124)
Observations	83346	83346	83346	83346
R-square	0.111	0.115	0.111	0.111
Y-mean	8.518	8.518	8.518	8.518
Y-sd	5.380	5.380	5.380	5.380
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S6.15: Placebo test using counties without plants in the same state

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	-0.135	-0.170	-0.132	-0.126
	(0.308)	(0.308)	(0.308)	(0.308)
PlannedDowntime	-0.144	-0.183	-0.141	-0.143
	(0.136)	(0.147)	(0.134)	(0.135)
UnplannedShut	-0.120	-0.066	-0.118	-0.118
	(0.257)	(0.255)	(0.256)	(0.256)
UnplannedDowntime	-0.128	-0.113	-0.125	-0.129
	(0.137)	(0.125)	(0.134)	(0.135)
OutageAnyPlant	-.0429	-.104	-.0407	-.0798
	(.129)	(.134)	(.13)	(.138)
Observations	110963	110963	110963	110963
R-square	0.127	0.132	0.127	0.128
Y-mean	8.382	8.382	8.382	8.382
Y-sd	5.506	5.506	5.506	5.506
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S6.16: Double difference analysis with control counties

	HCHO			
	(1)	(2)	(3)	(4)
Treated × PlannedShut	0.071	0.060	0.071	0.068
	(0.330)	(0.322)	(0.329)	(0.329)
Treated × PlannedDowntime	-0.289*	-0.297*	-0.289*	-0.289*
	(0.146)	(0.147)	(0.146)	(0.146)
Treated × UnplannedShut	0.668**	0.649**	0.668**	0.670**
	(0.249)	(0.245)	(0.248)	(0.248)
Treated × UnplannedDowntime	-0.067	-0.054	-0.067	-0.066
	(0.081)	(0.080)	(0.081)	(0.082)
PlannedShut	-.136	-.188	-.138	-.134
	(.236)	(.241)	(.236)	(.236)
PlannedDowntime	-.14	-.181	-.141	-.143
	(.172)	(.195)	(.172)	(.172)
UnplannedShut	-.105	-.0655	-.105	-.105
	(.209)	(.214)	(.209)	(.209)
UnplannedDowntime	-.119	-.112	-.119	-.122
	(.115)	(.114)	(.115)	(.116)
OutageAnyPlant	.0524	.0215	.0446	.017
	(.0753)	(.0815)	(.0778)	(.0765)
Observations	216298	216298	216298	216298
R-square	0.118	0.122	0.118	0.118
Y-mean	8.259	8.259	8.259	8.259
Y-sd	5.544	5.544	5.544	5.544
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

S7 Additional tables on local aggressive behaviors

S7.1 Tweets

Table S7.1: Top ten refinery plants and their surrounding tweets

	ExxonMobil Baytown	Flint Hills Resources Corpus Christi	WRB Refining Borger	Marathon Catlettsburg
	(1)	(2)	(3)	(4)
County	Harris, Texas	Nueces, Texas	Hutchinson, Texas	Boyd, Kentucky
Capacity (bpd)	584,000	350,000	146,000	306,000
Start year	1920	1981	1927	1916
#Tweets per day	3322.0	650.1	10.86	252.1
	[3947.9]	[921.6]	[20.27]	[392.3]
Sentiment	0.222	0.304	0.181	0.312
	[0.104]	[0.163]	[0.329]	[0.161]
#Health tweets	2.317	0.501	0.004	0.182
	[3.479]	[1.028]	[0.071]	[0.524]
#Pollution tweets	14.55	2.045	0.067	0.651
	[17.39]	[2.844]	[0.767]	[1.842]
#Offensive tweets	404.9	81.12	1.260	26.92
	[666.0]	[131.2]	[3.362]	[49.35]
#Racist tweets	65.02	8.304	0.135	2.463
	[115.5]	[14.40]	[0.567]	[5.311]
	Monroe Energy Trainer	Placid Refining Port Allen	ConocoPhillips Belle Chasse	Valero Benicia
	(5)	(6)	(7)	(8)
County	Delaware, Pennsylvania	West Baton Rouge, Louisiana	Plaquemines, Louisiana	Solano, California
Capacity	190,000	57,000	247,000	132,000
Start year	1912	1975	1971	1969
#Tweets per day	2790.9	1379.0	78.21	308.1
	[2198.5]	[1507.8]	[122.6]	[766.2]
Sentiment	0.073	0.134	0.237	0.226
	[0.016]	[0.091]	[0.225]	[0.106]
#Health tweets	1.641	0.796	0.0315	0.2355
	[2.032]	[1.440]	[0.1991]	[0.8310]
#Pollution tweets	10.23	5.877	0.1319	0.7029
	[15.47]	[6.365]	[0.4037]	[2.003]
#Offensive tweets	418.0	204.7	12.20	86.61
	[432.5]	[257.9]	[24.41]	[111.6]
#Racist tweets	54.61	33.62	1.937	9.543
	[74.70]	[44.00]	[4.431]	[14.26]
	BP Ferndale	HollyFrontier Tulsa		
	(9)	(10)		
County	Whatcom, Washington	Tulsa, Oklahoma		
Capacity	225,000	70,300		
Start year	1954	1913		
#Tweets per day	41.81	1761.8		
	[39.15]	[661.4]		
Sentiment	0.364	0.099		
	[0.171]	[0.017]		
#Health tweets	0.0397	1.114		
	[0.2596]	[1.235]		
#Pollution tweets	0.2360	7.493		
	[0.5625]	[4.738]		

#Offensive tweets	2.154	161.4
	[3.474]	[77.38]
#Racist tweets	0.066	14.05
	[0.301]	[11.24]

Notes: Standard deviations are reported in brackets.

Table S7.2: Correlation of sentiment score, pollution- and health- related tweets

	Sentiment score					
	(1)	(2)	(3)	(4)	(5)	(6)
Pollution dummy	-0.016***		- 0.016***	-0.015***	-0.010***	-0.010***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Health dummy		-0.150***	-0.150***	-0.152***	-0.149***	-0.151***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	25398177	25398177	25398177	25398177	25398177	25398177
R-square	0.000	0.000	0.000	0.003	0.005	0.008
Y-mean	0.071	0.071	0.071	0.071	0.071	0.071
Y-sd	0.372	0.372	0.372	0.372	0.372	0.372
Plant FEs				Y		Y
Year FEs					Y	Y
Month FEs					Y	Y
DOW FEs					Y	Y
Hour FEs					Y	Y

Table S7.3: Correlation of sentiment score, offensive and racist measures

	Panel A: Sentiment score					
	(1)	(2)	(3)	(4)	(5)	(6)
Offensive dummy	-0.817***		-0.763***	-0.762***	-0.760***	-0.759***
	(0.000)		(0.001)	(0.001)	(0.001)	(0.001)
Racism dummy		-5.518***	-0.883***	-0.833***	-0.800***	-0.761***
		(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	25378172	25378172	25378172	25378172	25378172	25378172
R-square	0.105	0.048	0.106	0.107	0.107	0.109
Y-mean	0.071	0.071	0.071	0.071	0.071	0.071
Y-sd	0.372	0.372	0.372	0.372	0.372	0.372
Plant FEs				Y		Y
Year FEs					Y	Y
Month FEs					Y	Y
DOW FEs					Y	Y
Hour FEs					Y	Y

Panel B: Racist dummy				
	(1)	(2)	(3)	(4)
Offensive dummy	0.061***	0.060***	0.060***	0.060***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	25378172	25378172	25378172	25378172
R-square	0.368	0.370	0.372	0.373
Y-mean	0.010	0.010	0.010	0.010
Y-sd	0.015	0.015	0.015	0.015
Plant FEs		Y		Y
Year FEs			Y	Y
Month FEs			Y	Y
DOW FEs			Y	Y
Hour FEs			Y	Y

Table S7.4: Effects of unplanned shutdowns on #all tweets

	#All tweets ($\times 10^3$)			
	(1)	(2)	(3)	(4)
UnplannedShut	-0.049	-0.031	-0.050	-0.059
	(0.141)	(0.137)	(0.141)	(0.138)
Downtime	0.208***	0.173***	0.207***	0.264***
	(0.042)	(0.041)	(0.042)	(0.041)
Observations	21910	21910	21910	21910
R-square	0.582	0.609	0.582	0.597
Y-mean	1.114	1.114	1.114	1.114
Y-sd	2.065	2.065	2.065	2.065
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S7.5: Effects of unplanned shutdowns on offensive and racist tweets, adding user FEs

	Offensive dummy ($\times 100$)		Racist dummy ($\times 100$)	
	(1)	(2)	(3)	(4)
UnplannedShut	1.814***	0.078	0.188**	0.022
	(0.665)	(0.137)	(0.057)	(0.045)
Downtime	-0.488	0.100	0.078	0.035**
	(0.248)	(0.069)	(0.042)	(0.010)

OutageAnyPlant	-0.510***	0.058	-0.076*	-0.008
	(0.090)	(0.075)	(0.032)	(0.039)
Observations	25398177	25398177	25398177	25398177
R-square	0.018	0.401	0.017	0.418
Y-mean	7.578	7.578	0.805	0.806
Y-sd	14.79	14.79	0.844	0.844
Plant FEs	Y		Y	
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
User FEs		Y		Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S7.6: Effects of unplanned shutdowns on offensive and racist tweets, dropping pollution-related tweets

	Panel A: #Offensive tweets			
	(1)	(2)	(3)	(4)
UnplannedShut	29.956*	35.289**	29.942*	33.548**
	(13.071)	(11.532)	(13.030)	(10.916)
Downtime	25.415	14.466	25.382	34.530
	(51.764)	(45.692)	(51.691)	(47.168)
OutageAnyPlant	-21.378**	-0.528	-21.564**	1.043
	(6.013)	(3.333)	(6.056)	(2.456)
Observations	10275	10275	10275	10275
R-square	0.477	0.507	0.477	0.488
Y-mean	137.389	137.389	137.389	137.389
Y-sd	303.943	303.943	303.943	303.943
	Panel B: #Racist tweets			
UnplannedShut	5.084*	5.822*	5.083*	5.610**
	(2.463)	(2.420)	(2.459)	(2.174)
Downtime	5.512	3.383	5.508	6.846
	(9.292)	(8.310)	(9.283)	(8.676)
OutageAnyPlant	-3.013**	0.147	-3.033**	0.272
	(0.921)	(0.509)	(0.941)	(0.407)
Observations	10275	10275	10275	10275
R-square	0.422	0.450	0.422	0.430
Y-mean	18.626	18.626	18.626	18.626
Y-sd	50.775	50.775	50.775	50.775
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S7.7: Summary statistics on offensive tweets against subgroups

	#Tweets	Proportion	
All	25,398,177		
Offensive	1,918,055	7.55%	
Racist	263,346	1.04%	
	Offensive against subgroups		
	#Tweets	Proportion out of all tweets	Proportion out of offensive tweets
Anti-government	168,951	0.67%	8.81%
Xenophobic	148,036	0.58%	7.72%
Sexual	801,183	3.15%	41.77%
Racist	231,030	0.91%	12.05%
Other	609,171	2.40%	31.76%

Notes: Racist in the lower panel denotes offensive and racist. The sum of four subgroups and other tweets is larger than #offensive tweets as tweets could be both anti-government and xenophobic, etc.

Table S7.8: Effects of unplanned shutdowns on offensive tweets against subgroups

	Offensive	Offensive against subgroups				
		Anti-government	Xenophobic	Sexual	Racist	Other
	(1)	(2)	(3)	(4)	(5)	(6)
UnplannedShut	33.560**	2.337***	3.105**	14.327***	3.748***	13.419
	(10.902)	(0.300)	(0.845)	(1.974)	(0.905)	(9.658)
Downtime	34.554	1.904	2.074	10.118	2.048	18.269
	(47.202)	(2.919)	(3.164)	(19.228)	(4.538)	(18.976)
OutageAnyPlant	1.013	0.219	0.450	1.874	0.539**	-1.009
	(2.465)	(0.345)	(0.233)	(1.071)	(0.153)	(1.157)
Observations	10275	10275	10275	10275	10275	10275
R-square	0.488	0.570	0.559	0.509	0.507	0.389
Y-mean	137.592	14.532	13.475	57.473	16.573	43.699
Y-sd	304.289	23.277	22.945	121.260	33.075	123.680
Plant FEs	Y	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S7.9: IV regression on tweet outcomes

	Panel A: Proportion of tweets with air pollution keywords (in percentage)			
	(1)	(2)	(3)	(4)
HCHO	0.108*	0.118**	0.109*	0.108*
	(0.058)	(0.056)	(0.058)	(0.058)
OutageAnyPlant	0.193	0.177	0.197	0.195
	(0.123)	(0.140)	(0.123)	(0.125)
Observations	8796	8796	8796	8796
R-square	-0.004	0.003	-0.005	-0.005
F-stat	30.822	29.336	30.753	30.754
Y-mean	0.551	0.551	0.551	0.551
Y-sd	1.857	1.857	1.857	1.857
	Panel B: Proportion of tweets with health keywords (in percentage)			
HCHO	0.002	0.003	0.002	0.002
	(0.011)	(0.011)	(0.011)	(0.011)
OutageAnyPlant	0.019	0.033	0.019	0.019
	(0.020)	(0.022)	(0.020)	(0.020)
Observations	8796	8796	8796	8796
R-square	0.019	0.025	0.019	0.019
F-stat	30.822	29.336	30.753	30.754
Y-mean	0.062	0.062	0.062	0.062
Y-sd	0.319	0.319	0.319	0.319
	Panel C: Sentiment			
HCHO	0.001	-0.000	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
OutageAnyPlant	-0.004	-0.004	-0.003	-0.007
	(0.006)	(0.006)	(0.006)	(0.006)
Observations	9833	9833	9833	9833
R-square	0.221	0.252	0.221	0.224
F-stat	28.537	27.827	28.493	28.535
Y-mean	0.120	0.120	0.120	0.120
Y-sd	0.108	0.108	0.108	0.108
	Panel D: #Offensive tweets			
HCHO	13.438**	16.038**	13.435**	15.048**
	(6.585)	(7.786)	(6.579)	(6.375)
OutageAnyPlant	-19.278***	2.209	-19.383***	3.475
	(4.812)	(6.911)	(4.939)	(4.471)
Observations	10275	10275	10275	10275
R-square	0.420	0.428	0.420	0.417
F-stat	10.544	12.654	10.557	10.449
Y-mean	137.592	137.592	137.592	137.592
Y-sd	304.289	304.289	304.289	304.289
	Panel E: #Racist tweets			
HCHO	2.284*	2.650*	2.283*	2.518**
	(1.303)	(1.517)	(1.303)	(1.284)
OutageAnyPlant	-2.624***	0.615	-2.631***	0.698
	(0.850)	(1.340)	(0.876)	(0.843)
Observations	10275	10275	10275	10275

R-square	0.363	0.373	0.363	0.359
F-stat	10.544	12.654	10.557	10.449
Y-mean	18.647	18.647	18.647	18.647
Y-sd	50.834	50.834	50.834	50.834
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S7.10: Effects of unplanned shutdowns on pollution tweets on lead and lag days

	Panel A: Proportion of tweets with air pollution keywords (in percentage) on lead day 1			
	(1)	(2)	(3)	(4)
UnplannedShut	0.182	0.187	0.184	0.184
	(0.237)	(0.237)	(0.237)	(0.237)
Downtime	-0.081	-0.025	-0.079	-0.088
	(0.073)	(0.074)	(0.073)	(0.073)
Observations	19237	19237	19237	19237
R-square	0.015	0.023	0.015	0.016
Y-mean	0.566	0.566	0.566	0.566
Y-sd	1.805	1.805	1.805	1.805
	Panel B: on lead day 2			
UnplannedShut	-0.134	-0.135	-0.132	-0.132
	(0.235)	(0.235)	(0.235)	(0.235)
Downtime	-0.074	-0.016	-0.072	-0.082
	(0.073)	(0.074)	(0.073)	(0.073)
Observations	19228	19228	19228	19228
R-square	0.015	0.023	0.015	0.015
Y-mean	0.566	0.566	0.566	0.566
Y-sd	1.806	1.806	1.806	1.806
	Panel C: on lag day 1			
UnplannedShut	-0.038	-0.028	-0.038	-0.038
	(0.235)	(0.235)	(0.235)	(0.235)
Downtime	-0.055	0.002	-0.055	-0.065
	(0.073)	(0.075)	(0.073)	(0.073)
Observations	19362	19362	19362	19362
R-square	0.009	0.013	0.009	0.009
Y-mean	0.567	0.567	0.567	0.567
Y-sd	1.810	1.810	1.810	1.810

	Panel D: #Offensive tweets on lead day 1			
UnplannedShut	2.099	9.809	2.098	1.844
	(11.217)	(13.666)	(11.260)	(14.154)
Downtime	26.996	16.648	26.995	35.552
	(45.965)	(40.623)	(45.919)	(41.938)
Observations	10275	10275	10275	10275
R-square	0.479	0.508	0.479	0.489
Y-mean	137.948	137.948	137.948	137.948
Y-sd	305.021	305.021	305.021	305.021
	Panel E: on lead day 2			
UnplannedShut	1.269	10.262	1.223	0.963
	(14.567)	(15.980)	(14.601)	(16.302)
Downtime	26.956	17.311	26.923	35.706
	(52.143)	(45.988)	(52.098)	(48.051)
Observations	10275	10275	10275	10275
R-square	0.478	0.507	0.478	0.489
Y-mean	138.449	138.449	138.449	138.449
Y-sd	307.212	307.212	307.212	307.212
	Panel F: on lag day 1			
UnplannedShut	2.044	9.497	2.013	1.748
	(12.638)	(14.982)	(12.667)	(14.519)
Downtime	25.228	16.093	25.205	34.148
	(48.446)	(43.673)	(48.389)	(44.042)
Observations	10275	10275	10275	10275
R-square	0.475	0.505	0.475	0.486
Y-mean	138.874	138.874	138.874	138.874
Y-sd	308.540	308.540	308.540	308.540
	Panel G: #Racist tweets on lead day 1			
UnplannedShut	-0.129	1.191	-0.130	-0.168
	(1.761)	(2.075)	(1.768)	(2.217)
Downtime	5.792	3.873	5.792	7.069
	(8.537)	(7.664)	(8.530)	(7.977)
Observations	10275	10275	10275	10275
R-square	0.423	0.452	0.423	0.432
Y-mean	18.673	18.673	18.673	18.673
Y-sd	50.935	50.935	50.935	50.935
	Panel H: on lead day 2			
UnplannedShut	0.330	1.782	0.323	0.284
	(2.773)	(2.954)	(2.776)	(2.972)
Downtime	5.703	3.886	5.698	7.011
	(9.262)	(8.304)	(9.255)	(8.702)
Observations	10275	10275	10275	10275
R-square	0.422	0.449	0.422	0.430
Y-sd	51.535	51.535	51.535	51.535
	Panel I: on lag day 1			
UnplannedShut	0.810	2.242	0.807	0.767
	(2.309)	(2.563)	(2.309)	(2.571)
Downtime	5.419	3.617	5.417	6.751

	(8.999)	(8.186)	(8.994)	(8.406)
Observations	10275	10275	10275	10275
R-square	0.421	0.449	0.421	0.430
Y-mean	18.830	18.830	18.830	18.830
Y-sd	51.556	51.556	51.556	51.556
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S7.11: Job concerns due to unplanned shutdowns

	#Tweets with job keywords
UnplannedShut	0.172
	(0.609)
Downtime	-0.022
	(0.074)
Observations	21233
R-square	0.043
Y-mean	4.085
Plant FEs	Y
Year FEs	Y
Month FEs	Y
DOW FEs	Y
Trends	Quadratic

Notes: Standard errors are clustered at the plant level. Job-related keywords include: job, jobless, employment, unemployment, nonemployment, employee, employer, career, work, occupation, profession, professional, vocation, application, income, tenure, wage, salary, payroll, paycheck, compensation, layoff, furlough, retirement, hire, recruit, contract, labor, labour, duty, task, workload, function, office, service.

S7.2 Crimes

Table S7.12: Effects of unplanned shutdowns on hate crimes against other groups

	Panel A: #Anti-White events			
	(1)	(2)	(3)	(4)
UnplannedShut	-0.003	-0.003	-0.003	-0.003
	(0.005)	(0.005)	(0.005)	(0.005)
Downtime	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	105335	105335	105335	105335
R-square	0.025	0.026	0.025	0.025
Y-mean	0.008	0.008	0.008	0.008
Y-sd	0.094	0.094	0.094	0.094
	Panel B: #Anti-Asian events			
UnplannedShut	-0.002	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Downtime	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	105335	105335	105335	105335
R-square	0.013	0.017	0.013	0.013
Y-mean	0.002	0.002	0.002	0.002
Y-sd	0.058	0.058	0.058	0.058
	Panel C: #Anti-Hispanic events			
UnplannedShut	-0.002	-0.002	-0.002	-0.002
	(0.007)	(0.007)	(0.007)	(0.007)
Downtime	0.000	0.000	-0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	105335	105335	105335	105335
R-square	0.082	0.085	0.082	0.082
Y-mean	0.012	0.012	0.012	0.012
Y-sd	0.119	0.119	0.119	0.119
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S7.13: IV regression on hate crimes

	Panel A: #Hate crime events against black people			
	(1)	(2)	(3)	(4)
HCHO	0.011***	0.009**	0.011***	0.011***
	(0.004)	(0.004)	(0.004)	(0.004)
OutageAnyPlant	0.010	0.013	0.010	0.011
	(0.008)	(0.010)	(0.008)	(0.008)
Observations	105335	105335	105335	105335
R-square	0.056	0.088	0.056	0.059
F-stat	12.774	16.097	12.616	12.959
Y-mean	0.031	0.031	0.031	0.031
Y-sd	0.192	0.192	0.192	0.192
	Panel B: #Black victims			
HCHO	0.008*	0.006	0.008*	0.008*
	(0.004)	(0.005)	(0.004)	(0.004)
OutageAnyPlant	0.016	0.016	0.015	0.015
	(0.014)	(0.015)	(0.014)	(0.014)
Observations	105335	105335	105335	105335
R-square	0.082	0.093	0.081	0.081
F-stat	12.774	16.097	12.616	12.959
Y-mean	0.034	0.034	0.034	0.034
Y-sd	0.259	0.259	0.259	0.259
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S7.14: Effects of unplanned shutdowns on hate crimes on lead and lag days

	Panel A: #Hate crime events against black people on lead day 1			
	(1)	(2)	(3)	(4)
UnplannedShut	-0.001	-0.000	-0.001	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
Downtime	0.003	0.003	0.003	0.003
	(0.003)	(0.002)	(0.003)	(0.003)
Observations	105335	105335	105335	105335
R-square	0.151	0.154	0.151	0.152
Y-mean	0.029	0.029	0.029	0.029
Y-sd	0.189	0.189	0.189	0.189
	Panel B: on lead day 2			
UnplannedShut	0.003	0.003	0.003	0.003
	(0.007)	(0.007)	(0.007)	(0.007)
Downtime	-0.000	0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)

Observations	105335	105335	105335	105335
R-square	0.144	0.147	0.144	0.144
Y-mean	0.029	0.029	0.029	0.029
Y-sd	0.185	0.185	0.185	0.185
	Panel C: on lag day 1			
UnplannedShut	-0.002	-0.001	-0.002	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Downtime	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	105335	105335	105335	105335
R-square	0.149	0.152	0.149	0.149
Y-mean	0.028	0.028	0.028	0.028
Y-sd	0.183	0.183	0.183	0.183
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S7.15: Results by crime types

	Homicide	Human Trafficking	Assault	Sex Offenses	Robbery
	(1)	(2)	(3)	(4)	(5)
UnplannedShut	0.217***	0.051*	1.264**	0.154	1.548***
	(0.036)	(0.023)	(0.566)	(0.277)	(0.491)
Downtime	0.060*	-0.084	-0.364	0.028	-0.844
	(0.032)	(0.047)	(0.642)	(0.068)	(0.561)
Observations	2234	432	16822	8455	10516
R-square	0.113	0.280	0.823	0.250	0.694
Y-mean	1.297	1.178	7.778	1.748	6.257
Estimate relative to mean	16.7%	4.3%	16.3%	8.8%	24.7%
	Burglary	Theft	Fraud	Drug/ Narcotic	Extortion/ Blackmail
UnplannedShut	1.917	0.513**	0.307	0.740**	0.283***
	(1.094)	(0.213)	(0.335)	(0.272)	(0.049)
Downtime	-0.314	-0.387	0.044	-0.175	-0.007
	(0.444)	(0.856)	(0.057)	(0.358)	(0.092)
Observations	27073	23891	2401	29648	1347
R-square	0.626	0.640	0.127	0.594	0.159
Y-mean	7.937	7.856	1.311	6.480	1.181
Estimate relative to mean	24.2%	6.5%	23.4%	11.4%	24.0%
County FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S7.16: IV regression on other crimes

	Panel A: #Crimes			
	(1)	(2)	(3)	(4)
HCHO	3.997	4.616	4.027	3.825
	(4.791)	(3.086)	(4.758)	(4.370)
OutageAnyPlant	-0.381	3.989	-0.527	1.695
	(0.924)	(2.579)	(0.976)	(1.548)
Observations	61831	61831	61831	61831
R-square	0.522	0.515	0.521	0.526
F-stat	15.173	8.939	14.599	15.790
Y-mean	42.963	42.963	42.963	42.963
Y-sd	107.560	107.560	107.560	107.560
	Panel B: #Victims			
HCHO	5.974	6.725	5.999	5.763
	(6.471)	(4.218)	(6.434)	(5.965)
OutageAnyPlant	-0.269	5.459	-0.407	2.287
	(1.178)	(3.556)	(1.242)	(2.024)
Observations	61831	61831	61831	61831
R-square	0.468	0.458	0.468	0.473
F-stat	15.173	8.939	14.599	15.790
Y-mean	49.906	49.906	49.906	49.906
Y-sd	128.149	128.149	128.149	128.149
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S7.17: Effects of unplanned shutdowns on other crimes on lead and lag days

	Panel A: #Crimes on lead day 1			
	(1)	(2)	(3)	(4)
UnplannedShut	0.888	0.734	0.894	1.404
	(1.341)	(1.305)	(1.335)	(1.219)
Downtime	-0.097	-0.297	-0.095	-0.048
	(1.019)	(1.217)	(1.018)	(0.877)
Observations	61818	61818	61818	61818
R-square	0.557	0.562	0.557	0.558
Y-mean	42.891	42.891	42.891	42.891
Y-sd	107.341	107.341	107.341	107.341
	Panel B: on lead day 2			
UnplannedShut	0.166	-0.074	0.174	0.646
	(2.034)	(2.134)	(2.027)	(2.025)
Downtime	-0.130	-0.339	-0.126	-0.083
	(0.967)	(1.097)	(0.966)	(0.833)

Observations	61773	61773	61773	61773
R-square	0.557	0.562	0.557	0.558
Y-mean	42.808	42.808	42.808	42.808
Y-sd	107.231	107.231	107.231	107.231
	Panel C: on lag day 1			
UnplannedShut	1.069	0.932	1.012	1.513
	(1.769)	(1.970)	(1.756)	(1.808)
Downtime	0.008	-0.165	-0.013	0.032
	(1.032)	(0.955)	(1.023)	(0.892)
Observations	61812	61812	61812	61812
R-square	0.558	0.564	0.558	0.559
Y-mean	42.758	42.758	42.758	42.758
Y-sd	106.982	106.982	106.982	106.982
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S7.18: Results using the union of hate crime reporting counties and non-hate crime reporting counties

	#Hate crimes				#Other crimes
	all victims	Black	Asian	Hispanic	
	(1)	(2)	(3)	(4)	(5)
UnplannedShut	0.011	0.007**	-0.002	-0.000	1.917**
	(0.010)	(0.003)	(0.001)	(0.002)	(0.666)
Downtime	-0.001	-0.002*	0.001	-0.000	1.417
	(0.002)	(0.001)	(0.001)	(0.001)	(0.972)
Observations	234602	234602	234602	234602	234602
R-square	0.291	0.146	0.011	0.078	0.568
Y-mean	0.109	0.029	0.002	0.011	45.204
County FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

S8 Additional tables on health

S8.1 Medical expenditure

Table S8.1: IV regression on medical expenditure

	Expenditure at the county-day level ($)		
	Cough remedies	Sinus remedies	Breathing aids
(Product module code)	(8425)	(8502)	(7790)
HCHO	0.318**	0.146**	0.070**
	(0.139)	(0.069)	(0.030)
OutageAnyPlant	0.183	0.061	0.000
	(0.129)	(0.060)	(0.002)
Observations	105335	105335	105335
R-square	0.033	-0.137	-0.272
F-stat	11.499	11.499	11.499
Y-mean	1.113	0.179	0.031
Y-sd	4.746	2.090	0.663
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.2: Effects of unplanned shutdowns on placebo product expenditure

	Expenditure at the county-day level ($)		
	Eyeglass accessories	Tooth & gum analgesics	Insoles
(Product module code)	(7925)	(8402)	(8441)
UnplannedShut	-0.049	-0.029	-0.078
	(0.062)	(0.049)	(0.116)
Downtime	0.023	-0.024*	0.014
	(0.018)	(0.014)	(0.033)
Observations	105335	105335	105335
R-square	0.012	0.035	0.054
Y-mean	0.084	0.093	0.251
Y-sd	1.087	0.869	2.079
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.3: Effects of unplanned shutdowns on purchases with high and negative income elasticity

	Expenditure at the county-day level ($)			
	Electronics, records and tapes	Automotive	Instant tea	Prepared food
	(1)	(2)	(3)	(4)
UnplannedShut	-0.012	-0.069	-0.001	-0.019
	(0.097)	(0.097)	(0.030)	(0.023)
Downtime	-0.119***	-0.112***	-0.002	0.018
	(0.028)	(0.034)	(0.005)	(0.017)
OutageAnyPlant	0.170**	-0.061	0.009	-0.006
	(0.064)	(0.053)	(0.012)	(0.015)
Observations	221291	221291	221291	221291
R-square	0.023	0.047	0.020	0.199
Y-mean	0.250	0.525	0.064	0.424
Y-sd	3.568	3.618	0.900	1.535
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.4: Effects of unplanned shutdowns on medical expenditure on lead and lag days

	Expenditure at the county-day level ($)		
	Panel A: on lead day 1		
	Cough remedies	Sinus remedies	Breathing aids
(Product module code)	(8425)	(8502)	(7790)
UnplannedShut	0.166	-0.061	-0.010
	(0.258)	(0.079)	(0.036)
Downtime	-0.014	-0.033	-0.001
	(0.074)	(0.023)	(0.010)
Observations	105335	105335	105335
R-square	0.142	0.040	0.012
Y-mean	1.104	0.176	0.035
Y-sd	4.861	1.417	0.631
	Panel B: on lead day 2		
UnplannedShut	0.321	-0.087	-0.034
	(0.248)	(0.080)	(0.040)
Downtime	-0.066	-0.003	-0.013
	(0.071)	(0.023)	(0.011)
Observations	105335	105335	105335
R-square	0.142	0.045	0.007
Y-mean	1.109	0.182	0.032
Y-sd	4.878	1.435	0.700

	Panel C: on lag day 1		
UnplannedShut	-0.036	-0.044	-0.002
	(0.259)	(0.082)	(0.034)
Downtime	-0.005	-0.045*	0.006
	(0.074)	(0.024)	(0.010)
Observations	105335	105335	105335
R-square	0.142	0.045	0.011
Y-mean	1.109	0.185	0.032
Y-sd	4.878	1.472	0.603
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.5: Effects of unplanned shutdowns on medical expenditure, only using households with complete data 2014-2019

	Expenditure at the county-day level ($)		
	Cough remedies	Sinus remedies	Breathing aids
(Product module code)	(8425)	(8502)	(7790)
UnplannedShut	0.605***	0.067***	0.009
	(0.198)	(0.023)	(0.023)
Downtime	0.039	-0.002	-0.004
	(0.057)	(0.006)	(0.006)
Observations	105335	105335	105335
R-square	0.106	0.006	0.005
Y-mean	0.683	0.091	0.016
Y-sd	3.663	1.000	0.455
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.6: Effects of unplanned shutdowns on medical expenditure by race

	Expenditure at the county-day level ($)		
	Cough remedies	Sinus remedies	Breathing aids
(Product module code)	(8425)	(8502)	(7790)
UnplannedShut	0.438***	0.019**	0.015*
	(0.167)	(0.010)	(0.009)
Black	-0.686***	-0.005***	-0.015***
	(0.013)	(0.001)	(0.001)
Black × UnplannedShut	-0.347	-0.019	-0.019
	(0.236)	(0.014)	(0.013)
Downtime	-0.008	0.000	-0.001
	(0.034)	(0.002)	(0.002)
Observations	210670	210670	210670
R-square	0.077	0.002	0.011
Y-mean	0.473	0.078	0.010
Y-sd	3.046	0.925	0.345
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.7: Population near refineries

	Nielsen consumer panel				
	within 25km	25-50km	50-75km	75-100km	>100km
	(1)	(2)	(3)	(4)	(5)
#Households surveyed	42.06	29.09	11.78	12.95	10.78
(per county)	[74.04]	[59.57]	[18.57]	[27.52]	[24.30]
White (%)	80.26	87.60	90.71	91.55	89.07
	[19.38]	[16.87]	[18.18]	[17.91]	[21.18]
Black (%)	11.23	6.14	3.84	3.92	7.15
	[15.94]	[10.16]	[11.73]	[12.70]	[18.55]
Asian (%)	3.24	2.19	1.35	2.27	0.95
	[6.33]	[5.08]	[4.52]	[9.70]	[4.24]
	American community survey				
Population surveyed	4182.9	4593.2	4751.1	4280.6	4445.6
(per census tract)	[2138.1]	[2487.1]	[2628.0]	[2169.9]	[2348.0]
White (%)	58.02	64.46	76.84	73.44	76.12
	[29.13]	[26.63]	[21.66]	[24.95]	[22.60]
Black (%)	20.87	13.49	7.78	12.87	12.98
	[28.87]	[20.83]	[14.83]	[22.00]	[19.96]
Asian (%)	8.67	9.42	5.99	4.55	3.14
	[13.56]	[13.14]	[11.16]	[8.83]	[5.55]

Notes: Standard deviations are reported in brackets.

S8.2 Foot traffic

Table S8.8: Effects of temporary planned and unplanned outages on HCHO, 2018-2019

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	-0.167	-0.152	-0.198	-0.204
	(0.688)	(0.683)	(0.691)	(0.691)
PlannedDowntime	-0.498*	-0.483*	-0.502*	-0.510**
	(0.240)	(0.238)	(0.244)	(0.240)
UnplannedShut	0.994*	0.988*	0.998*	1.003*
	(0.500)	(0.505)	(0.507)	(0.508)
UnplannedDowntime	-0.074	-0.103	-0.093	-0.085
	(0.236)	(0.229)	(0.239)	(0.239)
OutageAnyPlant	.156	.0702	.0366	.0463
	(.196)	(.305)	(.171)	(.172)
Observations	30425	30425	30425	30425
R-square	0.144	0.146	0.145	0.145
Y-mean	8.652	8.652	8.652	8.652
Y-sd	5.913	5.913	5.913	5.913
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.9: IV regression on foot traffic

	#Devices at the county-day level ($\times 10^5$)	#Visits at the county-day level ($\times 10^5$)	
		Amusement parks & recreational camps	General medical & surgical hospitals
	(1)	(2)	(3)
HCHO	0.280	0.037	1.977**
	(0.389)	(0.031)	(0.795)
OutageAnyPlant	0.026	-0.000	0.433
	(0.067)	(0.010)	(0.593)
Observations	30425	30425	30425
R-square	0.890	0.775	0.706
F-stat	11.764	11.764	11.764
Y-mean	3.075	0.350	11.942
Y-sd	5.602	0.599	26.452
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.10: Effects of unplanned shutdowns on foot traffic on lead and lag days

	#Devices at the county-day level ($\times 10^5$)	#Visits at the county-day level ($\times 10^5$)	
		Amusement parks & recreational camps	General medical & surgical hospitals
	Panel A: on lead day 1		
	(1)	(2)	(3)
UnplannedShut	0.080	-0.079	1.323
	(0.118)	(0.081)	(1.066)
Downtime	-0.157***	-0.053**	0.222
	(0.036)	(0.025)	(0.325)
Observations	30425	30425	30425
R-square	0.966	0.528	0.872
Y-mean	3.087	0.494	11.963
Y-sd	5.638	1.048	26.468
	Panel B: on lead day 2		
UnplannedShut	-0.030	-0.076	0.951
	(0.116)	(0.075)	(1.088)
Downtime	-0.155***	-0.055**	0.016
	(0.035)	(0.023)	(0.331)
Observations	30425	30425	30425
R-square	0.966	0.558	0.867
Y-mean	3.077	0.489	11.971
Y-sd	5.616	1.002	26.555
	Panel C: on lag day 1		
UnplannedShut	-0.118	-0.026	-0.423
	(0.119)	(0.073)	(1.073)
Downtime	-0.154***	-0.062***	-0.056
	(0.036)	(0.022)	(0.327)
Observations	30377	30425	30425
R-square	0.965	0.568	0.865
Y-mean	3.077	0.487	11.758
Y-sd	5.630	0.993	25.950
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

S8.3 Mortality

Table S8.11: IV regression on mortality

	Panel A: #Death at the county--day level ($\times 10^{-3}$)			
	(1)	(2)	(3)	(4)
HCHO	0.128	0.106	0.136	0.121
	(0.134)	(0.140)	(0.142)	(0.126)
OutageAnyPlant	-0.702	-0.666	-0.733	-0.662
	(0.728)	(0.691)	(0.760)	(0.687)
Observations	105335	105335	105335	105335
R-square	0.075	0.076	0.075	0.075
F-stat	12.774	16.097	12.616	12.959
Y-mean	0.911	0.911	0.911	0.911
Y-sd	33.458	33.458	33.458	33.458
	Panel B: Death proportion of sampled beneficiaries (in percentage $\times 10^{-3}$)			
HCHO	0.017	0.014	0.018	0.016
	(0.018)	(0.018)	(0.019)	(0.017)
OutageAnyPlant	-0.091	-0.086	-0.095	-0.086
	(0.095)	(0.090)	(0.099)	(0.089)
Observations	105335	105335	105335	105335
R-square	0.075	0.075	0.075	0.075
F-stat	12.774	16.097	12.616	12.959
Y-mean	0.118	0.118	0.118	0.118
Y-sd	4.341	4.341	4.341	4.341
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S8.12: Effects of unplanned shutdowns on mortality on lead and lag days

	Panel A: #Death at the county--day level ($\times 10^{-3}$) on lead day 1			
	(1)	(2)	(3)	(4)
UnplannedShut	-0.038	-0.051	-0.038	-0.040
	(0.046)	(0.085)	(0.047)	(0.049)
Downtime	-0.054	-0.075	-0.052	-0.049
	(0.064)	(0.078)	(0.062)	(0.059)
Observations	105335	105335	105335	105335
R-square	0.094	0.095	0.094	0.094
Y-mean	1.130	1.130	1.130	1.130
Y-sd	37.085	37.085	37.085	37.085

	Panel B: on lead day 2			
UnplannedShut	-0.047	-0.039	-0.047	-0.052
	(0.068)	(0.080)	(0.067)	(0.074)
Downtime	-0.179	-0.170	-0.181	-0.175
	(0.188)	(0.176)	(0.190)	(0.184)
Observations	105335	105335	105335	105335
R-square	0.090	0.090	0.090	0.090
Y-mean	1.120	1.120	1.120	1.120
Y-sd	37.720	37.720	37.720	37.720
	Panel C: on lag day 1			
UnplannedShut	-0.089	-0.144	-0.088	-0.096
	(0.107)	(0.161)	(0.108)	(0.119)
Downtime	-0.234	-0.265	-0.230	-0.221
	(0.249)	(0.274)	(0.244)	(0.234)
Observations	105335	105335	105335	105335
R-square	0.105	0.105	0.105	0.105
Y-mean	1.405	1.405	1.405	1.405
Y-sd	43.769	43.769	43.769	43.769
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

S9 Additional tables on spillovers

Table S9.1: Correlation of Facebook SCI and Twitter following relationship

	ln(SCI)		
	(1)	(2)	(3)
#Followers	0.002***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)
Observations	10426441	32290	32290
R-square	0.000	0.006	0.033
Y-mean	7.604	7.724	7.724
Y-sd	1.749	1.124	1.124
Followee county FEs			Y
Sample	All counties	Top ten plants' counties	Top ten plants' counties

Table S9.2: Correlation of followees' and followers' tweets

	Offensive		Racist	
	dummy	#tweets	dummy	#tweets
	(1)	(2)	(3)	(4)
Offensive	0.012**	0.002		
	(0.006)	(0.013)		
Racist			0.001**	0.009
			(0.000)	(0.006)
Observations	5878453	5878453	5878453	5878453
R-square	0.070	0.001	0.000	0.029
Y-mean	0.015	0.026	0.001	0.002
Y-sd	0.120	0.353	0.037	0.054
Follower FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the follower level.

Table S9.3: Effects of unplanned shutdowns on followers' tweets

	Panel A			
	Offensive ($\times 10^2$)		Racist ($\times 10^2$)	
	dummy	#tweets	dummy	#tweets
	(1)	(2)	(3)	(4)
UnplannedShut	0.332	0.630	0.028	0.073
	(0.226)	(1.062)	(0.025)	(0.060)
Downtime	-0.070	-0.639	-0.037	-0.029
	(0.096)	(0.614)	(0.026)	(0.048)
Observations	5878453	5878453	5878453	5878453
R-square	0.079	0.019	0.032	0.028
Y-mean	2.062	5.420	0.135	0.178
Y-sd	15.594	57.29	4.025	6.354
	Panel B			
	Offensive ($\times 10^2$)		Racist ($\times 10^2$)	
	dummy	#tweets	dummy	#tweets
UnplannedShut	0.332**	0.630**	0.028*	0.073***
	(0.133)	(0.317)	(0.017)	(0.027)
Downtime	-0.070	-0.639***	-0.037	-0.029
	(0.051)	(0.120)	(0.025)	(0.044)
Observations	5878453	5878453	5878453	5878453
R-square	0.079	0.019	0.032	0.028
Y-mean	2.062	5.420	0.135	0.178
Y-sd	15.594	57.29	4.025	6.354
	Panel C			
	Offensive ($\times 10^2$)		Racist ($\times 10^2$)	
	dummy	#tweets	dummy	#tweets
UnplannedShut	0.332***	0.630	0.028***	0.073***
	(0.120)	(1.693)	(0.009)	(0.014)
Downtime	-0.070**	-0.639***	-0.037	-0.029
	(0.033)	(0.046)	(0.032)	(0.050)
Observations	5878453	5878453	5878453	5878453
R-square	0.079	0.019	0.032	0.028
Y-mean	2.062	5.420	0.135	0.178
Y-sd	15.594	57.29	4.025	6.354
Follower FEs	Y	Y	Y	Y
Plant FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors in Panel A are clustered at the follower level, 2,683 clusters in total. Standard errors in Panel B are clustered at the followee level, 920 clusters in total. Standard errors in Panel C are not clustered.

Table S9.4: Effects of unplanned shutdowns on distant air pollution

	HCHO			
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.226	0.001	6.446	-0.239
	(0.560)	(0.001)	(5.625)	(0.765)
OutageAnyPlant	-0.282	-0.105*	-0.243*	-0.089
	(0.444)	(0.054)	(0.131)	(0.054)
Days	0.668	0.679	0.599	0.646
	(2.284)	(2.284)	(2.272)	(2.291)
$Days^2$	-0.755**	-0.758**	-0.752**	-0.756**
	(0.317)	(0.316)	(0.314)	(0.315)
Observations	38679	38679	38679	38679
R-square	0.098	0.098	0.098	0.098
Y-mean	8.149	8.149	8.149	8.149
Y-sd	5.277	5.277	5.277	5.277
$Connected_{ij}$	ln(SCI)	Relative SCI	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Simple mean	Weighted mean
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic
Distance cutoff	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.5: Effects of unplanned shutdowns on geographically distant tweets, controlling for pollution dispersion

	#Offensive tweets		#Racist tweets	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***	0.098***	0.022***	0.014***
	(0.022)	(0.015)	(0.003)	(0.002)
OutageAnyPlant	-0.127***	-0.082***	-0.017***	-0.012***
	(0.017)	(0.012)	(0.002)	(0.001)
UnplannedShut × Dispersion		0.007***		0.001***
		(0.002)		(0.000)
Days	-0.480***	-0.476***	-0.050***	-0.049***
	(0.043)	(0.042)	(0.006)	(0.005)
$Days^2$	.145***	.145***	.0174***	.0174***
	(.0113)	(.0112)	(.00143)	(.00143)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.030	0.013	0.013
Y-mean	4.992	4.992	1.805	1.805
Y-sd	8.509	8.509	1.804	1.804

Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.6: Effects of unplanned shutdowns on geographically distant tweets, controlling for county-to-county migration flow

	#Offensive tweets		#Racist tweets	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***	0.162***	0.022***	0.022***
	(0.022)	(0.022)	(0.003)	(0.003)
OutageAnyPlant	-0.127***	-0.126***	-0.017***	-0.017***
	(0.017)	(0.017)	(0.002)	(0.002)
UnplannedShut × Inflow		0.008**		0.001**
		(0.004)		(0.000)
Days	-0.480***	-0.479***	-0.050***	-0.050***
	(0.043)	(0.043)	(0.006)	(0.006)
Days^2	.145***	.145***	.0174***	.0174***
	(.0113)	(.0113)	(.00143)	(.00143)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.028	0.013	0.013
Y-mean	4.992	4.992	1.805	1.805
Y-sd	8.509	8.509	1.804	1.804
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.7: Effects of unplanned shutdowns on geographically distant tweets, controlling for colocation probability

	#Offensive tweets		#Racist tweets	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***	0.071***	0.022***	0.012***
	(0.022)	(0.021)	(0.003)	(0.003)
OutageAnyPlant	-0.127***	-0.108***	-0.017***	-0.015***
	(0.017)	(0.015)	(0.002)	(0.002)
UnplannedShut × Colocation		0.006***		0.001***
		(0.001)		(0.000)
Days	-0.480***	-0.436***	-0.050***	-0.045***
	(0.043)	(0.036)	(0.006)	(0.005)
$Days^2$	.145***	.144***	.0174***	.0173***
	(.0113)	(.0112)	(.00143)	(.00142)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.029	0.013	0.013
Y-mean	4.992	4.992	1.805	1.805
Y-sd	8.509	8.509	1.804	1.804
$Connected_{ij}$	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.8: Effects of unplanned shutdowns on geographically distant tweets, controlling for geographic distance

	#Offensive tweets		#Racist tweets	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***	0.162***	0.022***	0.022***
	(0.022)	(0.022)	(0.003)	(0.003)
OutageAnyPlant	-0.127***	-0.127***	-0.017***	-0.017***
	(0.017)	(0.017)	(0.002)	(0.002)
UnplannedShut × Geodistance		0.036*		0.006**
		(0.019)		(0.003)
Days	-0.480***	-0.480***	-0.050***	-0.050***
	(0.043)	(0.043)	(0.006)	(0.006)
$Days^2$	.145***	.145***	.0174***	.0174***
	(.0113)	(.0113)	(.00143)	(.00143)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.028	0.013	0.013
Y-mean	4.992	4.992	1.805	1.805
Y-sd	8.509	8.509	1.804	1.804

Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150

Notes: The unit of *Geodistance* is $km \times 10^{-3}$. Standard errors are clustered at the county level.

Table S9.9: Effects of unplanned shutdowns on geographically distant tweets, controlling for traditional media subscription

	#Offensive tweets		#Racist tweets	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***	0.159***	0.022***	0.022***
	(0.022)	(0.021)	(0.003)	(0.003)
OutageAnyPlant	-0.127***	-0.124***	-0.017***	-0.017***
	(0.017)	(0.017)	(0.002)	(0.002)
UnplannedShut × Circulation		0.006		0.001*
($\times 10^{-2}$)		(0.005)		(0.000)
Days	-0.480***	-0.480***	-0.050***	-0.050***
	(0.043)	(0.043)	(0.006)	(0.006)
Days^2	.145***	.145***	.0174***	.0174***
	(.0113)	(.0112)	(.00143)	(.00143)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.029	0.013	0.013
Y-mean	4.992	4.992	1.805	1.805
Y-sd	8.509	8.509	1.804	1.804
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.10: Effects of unplanned shutdowns on geographically distant tweets, offensive against subgroups

	Offensive	Offensive against subgroups				
		Anti-government	Xenophobic	Sexual	Racist	Other
	(1)	(2)	(3)	(4)	(5)	(6)
UnplannedShut × Connected	0.163***	0.022***	0.022***	0.095***	0.030***	0.036***
	(0.022)	(0.004)	(0.003)	(0.015)	(0.005)	(0.003)
OutageAnyPlant	-0.127***	-0.017***	-0.017***	-0.074***	-0.024***	-0.028***
	(0.017)	(0.003)	(0.003)	(0.011)	(0.004)	(0.002)
Observations	5002053	5002053	5002053	5002053	5002053	5002053
R-square	0.028	0.010	0.009	0.017	0.012	0.024
Y-mean	4.850	0.679	0.648	2.767	0.857	1.219
Y-sd	19.458	5.363	5.465	10.239	6.421	7.624
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max	Max	Max
County FEs	Y	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Distance cutoff	150	150	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.11: Effects of unplanned shutdowns on distant tweets, alternative connectedness and aggregation

	Sentiment	#Tweets ($\times 10^{-2}$) with			
		Pollution	Health	Offensive	Racist content
	(1)	(2)	(3)	(4)	(5)
UnplannedShut × Connected	0.000	0.180	0.008	0.002	0.000
	(0.000)	(0.135)	(0.008)	(0.001)	(0.000)
R-square	0.021	0.003	0.001	0.032	0.014
Connected_{ij}	Relative SCI	Relative SCI	Relative SCI	Relative SCI	Relative SCI
Aggregation	Max	Max	Max	Max	Max
			Panel B		
UnplannedShut × Connected	0.048	-1.947	-1.730	0.115**	0.023***
	(0.081)	(16.782)	(1.645)	(0.056)	(0.009)
R-square	0.020	0.002	0.001	0.029	0.013
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Simple mean	Simple mean	Simple mean	Simple mean	Simple mean
			Panel C		
UnplannedShut × Connected	0.001	0.499	-0.195	0.200**	0.036*
	(0.013)	(0.878)	(0.260)	(0.090)	(0.021)
R-square	0.021	0.004	0.001	0.036	0.017
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Weighted mean	Weighted mean	Weighted mean	Weighted mean	Weighted mean
County FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Distance cutoff	150	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.12: Effects of unplanned shutdowns on distant tweets, alternative distance cutoffs

	Sentiment	#Tweets ($\times 10^{-2}$) with			
		Pollution	Health	Offensive	Racist content
	(1)	(2)	(3)	(4)	(5)
UnplannedShut × Connected	-0.017	-1.599	-0.168	0.141***	0.020***
	(0.022)	(1.510)	(0.470)	(0.017)	(0.002)
Observations	777969	5865307	5865307	5865307	5865307
R-square	0.018	0.001	0.001	0.031	0.013
Y-mean	0.105	2.412	0.363	5.100	1.218
Y-sd	0.302	56.534	19.864	20.024	9.952
Distance cutoff	100	100	100	100	100
			Panel B		
UnplannedShut × Connected	-0.003	-1.561	0.147	0.180***	0.025***
	(0.015)	(2.078)	(0.668)	(0.027)	(0.004)
Observations	527266	4088406	4088406	4088406	4088406
R-square	0.021	0.001	0.001	0.026	0.013
Y-mean	0.108	2.440	0.352	4.846	0.998
Y-sd	0.301	58.083	19.499	21.548	10.016
Distance cutoff	200	200	200	200	200
County FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max	Max

Notes: Standard errors are clustered at the county level.

Table S9.13: Effects of unplanned shutdowns on distant tweets, dropping retweeted tweets with comments

	#Tweets ($\times 10^{-2}$) with	
	Offensive	Racist content
	(1)	(2)
UnplannedShut × Connected	0.162***	0.022***
	(0.021)	(0.003)
Observations	5002053	5002053
R-square	0.028	0.013
Y-mean	4.945	1.016
Y-sd	19.053	9.662
Distance cutoff	150	150
County, year FEs	Y	Y
Month, DOW FEs	Y	Y
Trends	Quadratic	Quadratic
Connected_{ij}	ln(SCI)	ln(SCI)
Aggregation	Max	Max

Notes: Standard errors are clustered at the county level.

Table S9.14: Results using NIBRS and municipal 911 records

	#Non-hate crimes in distant counties replacing NIBRS with 911	
	(1)	(2)
UnplannedShut × Connected	0.344***	0.344***
	(0.083)	(0.086)
Observations	3065209	3065209
R-square	0.152	0.191
Y-mean	7.983	7.983
Y-sd	33.064	33.064
County FEs	Y	
Year FEs	Y	
County-year FEs		Y
DOW FEs	Y	Y
Month FEs	Y	Y
Trends	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.15: Effects of unplanned shutdowns on oil prices

	Spot prices		Futures prices	Retail prices
	Conventional gasoline	RBOB regular gasoline		
	(1)	(2)	(3)	(4)
UnplannedShut	-0.198	-0.003	-0.225	-0.014*
	(0.270)	(0.007)	(0.322)	(0.008)
Downtime	0.333*	0.006	0.191	0.023*
	(0.184)	(0.004)	(0.183)	(0.012)
OutageAnyPlant	1.616***	0.082***	2.316***	0.084***
	(0.003)	(0.000)	(0.003)	(0.005)
Observations	152914	152914	151298	179861
R-square	0.879	0.846	0.877	0.880
Y-mean	64.192	1.823	59.826	2.730
Y-sd	19.809	0.468	18.033	0.551
Plant FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S9.16: Effects of unplanned shutdowns on distant tweets, adding oil prices as controls

	#Tweets ($\times 10^{-2}$) with offensive content				
	(1)	(2)	(3)	(4)	(5)
UnplannedShut × Connected	0.163***	0.151***	0.154***	0.152***	0.169***
	(0.022)	(0.021)	(0.022)	(0.021)	(0.022)
Oil price		0.025***	0.031***	0.029***	0.022
		(0.002)	(0.003)	(0.003)	(0.022)
Observations	5002053	3433632	3490707	3424500	4992921
R-square	0.028	0.030	0.030	0.028	0.030
Y-mean	4.850	5.050	4.922	4.949	4.828
Oil price		Spot conventional	Spot RBOB regular	Future	Retail
County FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Connected$_{ij}$	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max	Max
Distance cutoff	150	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.17: Effects of unplanned shutdowns on geographically distant tweets, recentered treatment

	Panel A: Reshuffle at the time level			
	#Offensive tweets		#Racist tweets	
	Original	Recentered	Original	Recentered
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***		0.022***	
	(0.022)		(0.003)	
Recentered		0.095***		0.016***
		(0.015)		(0.005)
OutageAnyPlant	-0.127***	-0.118***	-0.017***	-0.013**
	(0.017)	(0.016)	(0.002)	(0.006)
Days	-0.480***	-15.957***	-0.050***	-2.514***
	(0.043)	(1.433)	(0.006)	(0.276)
Days^2	.145***	4.75***	.0174***	.864***
	(.0113)	(.37)	(.00143)	(.0711)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.028	0.013	0.013
Y-mean	4.850	4.850	1.538	1.538
Y-sd	19.458	19.458	6.159	6.159

	Panel B: Reshuffle at both plant and time level			
	#Offensive tweets		#Racist tweets	
	Original	Recentered	Original	Recentered
UnplannedShut × Connected	0.163***		0.022***	
	(0.022)		(0.003)	
Recentered		0.096***		0.013***
		(0.013)		(0.002)
OutageAnyPlant	-0.127***	-0.076***	-0.017***	-0.010***
	(0.017)	(0.010)	(0.002)	(0.001)
Days	-0.480***	-0.471***	-0.050***	-0.049***
	(0.043)	(0.042)	(0.006)	(0.005)
$Days^2$	.145***	.145***	.0174***	.0174***
	(.0113)	(.0113)	(.00143)	(.00144)
Observations	5002053	4999862	5002053	4999862
R-square	0.028	0.028	0.013	0.013
Y-mean	4.850	4.850	1.538	1.538
Y-sd	19.458	19.458	6.159	6.159
$Connected_{ij}$	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150

Notes: Standard errors are clustered at the county level.

Table S9.18: Local tweet effects with reverse spillovers

	Panel A: #Offensive tweets			
	(1)	(2)	(3)	(4)
UnplannedShut	29.237*	34.249**	29.225*	33.055**
	(13.213)	(11.404)	(13.172)	(10.978)
Downtime	25.369	14.379	25.337	34.512
	(51.802)	(45.730)	(51.729)	(47.201)
UnplannedShut_Distant × Connected	6.368	9.551**	6.355	4.421
	(4.839)	(3.728)	(4.843)	(3.623)
OutageAnyPlant	-21.816***	-1.117	-21.998***	0.746
	(5.868)	(3.415)	(5.914)	(2.510)
Observations	10275	10275	10275	10275
R-square	0.477	0.507	0.477	0.488
Y-mean	137.592	137.592	137.592	137.592
Y-sd	304.289	304.289	304.289	304.289

	Panel B: #Racist tweets			
UnplannedShut	4.889*	5.571*	4.887*	5.447**
	(2.432)	(2.357)	(2.428)	(2.133)
Downtime	5.492	3.354	5.488	6.830
	(9.311)	(8.326)	(9.302)	(8.693)
UnplannedShut_Distant × Connected	1.757*	2.359**	1.755*	1.473*
	(0.736)	(0.839)	(0.736)	(0.629)
OutageAnyPlant	-3.122**	0.013	-3.141**	0.184
	(0.911)	(0.531)	(0.931)	(0.420)
Observations	10275	10275	10275	10275
R-square	0.422	0.450	0.422	0.430
Y-mean	18.647	18.647	18.647	18.647
Y-sd	50.834	50.834	50.834	50.834
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S9.19: Local tweet effects in online isolated counties

	Panel A: #Offensive tweets			
	(1)	(2)	(3)	(4)
UnplannedShut	41.694***	39.124***	41.706***	38.173***
	(4.135)	(10.060)	(4.149)	(2.348)
Downtime	-20.594***	-6.267***	-20.591***	-8.474***
	(4.412)	(0.166)	(4.419)	(2.663)
OutageAnyPlant	-5.948	9.326	-5.925	-0.518
	(4.591)	(6.572)	(4.608)	(0.233)
Observations	2785	2785	2785	2785
R-square	0.527	0.151	0.527	0.540
Y-mean	143.379	143.379	143.379	143.379
Y-sd	291.913	291.913	291.913	291.913
	Panel B: #Racist tweets			
UnplannedShut	41.694***	39.124***	41.706***	38.173***
	(4.135)	(10.060)	(4.149)	(2.348)
Downtime	-20.594***	-6.267***	-20.591***	-8.474***
	(4.412)	(0.166)	(4.419)	(2.663)
OutageAnyPlant	-5.948	9.326	-5.925	-0.518
	(4.591)	(6.572)	(4.608)	(0.233)
Observations	2785	2785	2785	2785
R-square	0.527	0.151	0.527	0.540
Y-mean	143.379	143.379	143.379	143.379
Y-sd	291.913	291.913	291.913	291.913

Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S9.20: Local non-hate crime effects with reverse spillovers

	#Crimes			
	(1)	(2)	(3)	(4)
UnplannedShut	4.832**	5.386*	4.827**	4.923**
	(2.183)	(2.786)	(2.194)	(2.166)
Downtime	-0.621	-0.948	-0.641	-0.592
	(1.165)	(1.066)	(1.156)	(1.018)
UnplannedShut_Distant × Connected	1.761*	2.243*	1.750*	1.721
	(1.006)	(1.227)	(0.997)	(1.007)
OutageAnyPlant	-0.780	2.924	-0.997	0.963
	(0.776)	(1.676)	(0.785)	(1.076)
Observations	61831	61831	61831	61831
R-square	0.557	0.562	0.557	0.558
Y-mean	42.963	42.963	42.963	42.963
Y-sd	107.560	107.560	107.560	107.560
County, DOW FEs	Y	Y	Y	Y
Year, month FEs	Y		Y	Y
Year-month FEs		Y		
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.21: Local non-hate crime effects in online isolated counties

	#Crimes			
	(1)	(2)	(3)	(4)
UnplannedShut	7.920**	8.804***	7.816**	7.615**
	(2.612)	(2.172)	(2.661)	(2.586)
Downtime	-7.410	-4.928	-7.435	-7.348
	(6.872)	(4.575)	(6.865)	(6.546)
OutageAnyPlant	-2.814	5.033	-3.017	0.090
	(3.011)	(3.428)	(3.037)	(2.684)
Observations	17854	17854	17854	17854
R-square	0.486	0.514	0.487	0.490
Y-mean	39.758	39.758	39.758	39.758
Y-sd	82.572	82.572	82.572	82.572
County, DOW FEs	Y	Y	Y	Y
Year, month FEs	Y		Y	Y
Year-month FEs		Y		
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.22: Results with county-year fixed effects

	Panel A: #Hate crimes against black people	
	(1)	(2)
UnplannedShut	0.017**	0.018**
	(0.008)	(0.008)
Downtime	-0.002**	-0.003**
	(0.001)	(0.002)
Observations	105335	105335
R-square	0.151	0.155
	Panel B: #Non-hate crimes in local counties	
UnplannedShut	5.179**	5.273***
	(2.289)	(2.294)
Downtime	-0.593	-0.454
	(1.016)	(1.540)
Observations	61831	61831
R-square	0.558	0.812
	Panel C: #Non-hate crimes in distant counties	
UnplannedShut × Connected	0.319***	0.322***
	(0.084)	(0.087)
Observations	3060827	3060827
R-square	0.082	0.086
County FEs	Y	
Year FEs	Y	
County-year FEs		Y
DOW FEs	Y	Y
Month FEs	Y	Y
Trends	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.23: Remote tweet effects with connected plants' downtime controlled

	#Offensive tweets		#Racist tweets	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.163***	0.163***	0.022***	0.022***
	(0.022)	(0.022)	(0.003)	(0.003)
Downtime × Connected		-0.000		0.000
		(0.001)		(0.000)
Observations	5002053	5002053	5002053	5002053
R-square	0.028	0.028	0.013	0.013
Y-mean	4.850	4.850	1.538	1.538
Y-sd	19.458	19.458	6.159	6.159
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.24: Remote crime effects with connected plants' downtime controlled

	#Other crimes	
	(1)	(2)
UnplannedShut × Connected	0.319***	0.306***
	(0.084)	(0.095)
Downtime × Connected		0.061
		(0.155)
Observations	3060827	3060827
R-square	0.082	0.086
Y-mean	7.126	7.126
Y-sd	21.298	21.298
Connected_{ij}	ln(SCI)	ln(SCI)
Aggregation	Max	Max
County FEs	Y	Y
Year FEs	Y	Y
Month FEs	Y	Y
DOW FEs	Y	Y
Distance cutoff	150	150
Trends	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.25: Results on remote crimes using the union of hate crime reporting counties and non-hate crime reporting counties

	#Hate crimes ($\times 10^{-4}$)			#Other crimes
	all victims	Black	Asian	
	(1)	(2)	(3)	(4)
UnplannedShut × Connected	0.364	4.630	-0.949	0.265***
	(6.968)	(3.539)	(0.773)	(0.072)
Observations	3737846	3737846	3737846	3737846
R-square	0.163	0.054	0.007	0.099
Y-mean	53.266	15.838	0.993	5.836
Y-sd	823.235	413.486	102.533	19.467
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.26: Dropping counties with small Black population sizes

	#Hate crimes against Black people ($\times 10^{-4}$)		
	all remote	dropping bottom	only top quartile
	(1)	(2)	(3)
UnplannedShut × Connected	8.283	9.788	36.124*
	(6.468)	(7.898)	(21.285)
Observations	2381617	2048585	762468
R-square	0.054	0.054	0.067
Y-mean	24.857	28.278	57.983
Y-sd	517.791	552.376	797.009
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Distance cutoff	150	150	150
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.27: Results on tweets separating highly and loosely connected counties

	#Offensive tweets		#Racist tweets	
	Top quartile	Bottom	Top	Bottom
UnplannedShut × Connected	0.147***	0.002	0.019***	0.000
	(0.035)	(0.005)	(0.005)	(0.001)
Observations	1248870	1251061	1248870	1251061
R-square	0.066	0.033	0.028	0.007
Y-mean	4.371	3.424	1.570	1.391
Y-sd	32.440	39.057	6.699	6.014
Connected_{ij}	ln(SCI)	ln(SCI)	ln(SCI)	ln(SCI)
Aggregation	Max	Max	Max	Max
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Distance cutoff	150	150	150	150
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S9.28: Results on crimes separating highly and loosely connected counties

	#Other crimes	
	Top quartile	Bottom quartile
UnplannedShut × Connected	0.236**	-0.013
	(0.113)	(0.012)
Observations	1240106	578424
R-square	0.142	0.578
Y-mean	11.576	3.102
Y-sd	30.369	9.284
Connected_{ij}	ln(SCI)	ln(SCI)
Aggregation	Max	Max
County FEs	Y	Y
Year FEs	Y	Y
Month FEs	Y	Y
DOW FEs	Y	Y
Distance cutoff	150	150
Trends	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

S10 Results on mechanisms

Table S10.1: Effects of unplanned shutdowns on impulse control measures

	Expenditure on ($)			#Visits to
	Desserts	Cookies	Ice cream	gambling industries
(Product module code)	(1008)	(1505)	(2005)	
UnplannedShut	0.284**	0.414	0.612*	1.190
	(0.135)	(0.424)	(0.358)	(1.302)
Downtime	0.010	-0.029	-0.181	0.850
	(0.042)	(0.131)	(0.143)	(0.871)
Observations	105335	105335	105335	30425
R-square	0.379	0.686	0.664	0.340
Y-mean	1.687	7.185	8.546	5.317
Y-sd	4.276	14.793	18.827	31.631
County FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S10.2: Effects of unplanned shutdowns on police governance

	#Officers	#Tickets
	(1)	(2)
UnplannedShut	0.173	-14.701
	(0.204)	(9.307)
Downtime	-0.328	-16.624
	(0.211)	(12.060)
Observations	105335	105335
R-square	0.559	0.496
Y-mean	3.043	52.559
Y-sd	7.990	225.844
County FEs	Y	Y
Year FEs	Y	Y
Month FEs	Y	Y
DOW FEs	Y	Y
Trends	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

S11 Potential policy solutions

In this section, I assess current environmental regulations in the U.S. and Texas on refinery pollution and highlight their limitations. I explore a policy alternative that could reduce the adverse impacts of abnormal refinery outages. Current regulations of self reporting are not sufficient to capture plants' excess emissions, and the underreporting is increasingly severe over time. Utilization rates are positively correlated with the frequency and the per event pollution increase of abnormal operations. Using a cost-benefit analysis, I show lowering refinery plants' utilization rates generates net benefits and could improve social welfare.

S11.1 Self reporting in Texas

The U.S. Environmental Protection Agency (EPA) made the first Petroleum Refinery maximum achievable control technology standard in August 1995. The rule, known as 'Refinery MACT 1', covers all emission sources from petroleum refinery process units (except those regulated by other MACT standards). The rule specifies emissions standards of hazardous air pollutants (HAP) during refineries' normal operations. After decades of implementation, the EPA proposed a revised standard in December 2015 and planned to make it effective in February 2016. The revised standard requires 'refineries to meet the leak detection and repair' and states 'the need for alternative standards during startup and shutdown situations'. However, the implementation of the new regulation was suspended till now. There is no federal regulation on refinery's abnormal operations including excess emissions when unexpected outages.

The only exception is Texas which leads all states in energy resource endowment and crude oil production. It requires the self reporting of excess emission events from all industrial facilities including refineries since 1990. Information on shutdowns and operational changes must be provided to the Texas Commission on Environmental Quality (TCEQ) (EIA, 2007). Among all the excess emission events, refineries' emissions account for 14% during 2014-2019.[97] Facilities submit emission disclosure forms within 24 hours of the incidence. After the submission, the TCEQ posts emission events publicly available online 1-2 weeks later. As the policy relies on facilities' self reporting, three concerns exist about refinery emissions. First, to what degree do refineries report their abnormal emission events? Without sufficient ambient air quality or end-of-pipeline emission monitors, refinery plants have incentives to underreport their emission occurrence. Second, excess emission events refer to those over pollutant permit limits (in pounds per hour). Even if refineries report all excess events, are those unplanned outages generating pollutants under permit limit also harmful for surrounding air quality and social outcomes? Third, the policy is all about reporting to the regulator. We need standards or penalties to lower the emissions and their impacts, as well as broad public dissemination of the pollution information.

[97] 14% is calculated based on the Air Emission Event Report Database from the TCEQ. Data could be downloaded here: https://www2.tceq.texas.gov/oce/eer/

I answer the first question by estimating the underreporting rate. I use detailed schedules of plants' actual outages and reported emission events from the TCEQ. The estimation equation is as follows:

$$\#Reported_{it} = \beta_1 \#Unplanned_{it} + Time_t + Plant_i + \varepsilon_{it} \tag{9}$$

where the sample covers refinery plants in Texas at the plant-month level 2014-2019. $\#Reported_{it}$ is the number of reported excess emissions to the TCEQ by plant i in month t. $\#Unplanned_{it}$ is the number of actual unplanned outage events of plant i in month t. The coefficient of interest is β_1, measuring the proportion of unplanned outages reported to the regulator. I also add plant, year and month fixed effects as control variables.

The estimation result is shown in Table S11.1. As the number of unplanned outage increases by 1, reported event increases by 0.35, suggesting only 35% unplanned outages are captured in the excess emission records. In Column (1)-(3), the coefficient stays stable with plant, year and month fixed effects added. The finding shows only one third of unplanned outages are reflected to the regulator and disclosed on the website. A larger proportion of outages are not reported or available to the public. I also separately estimate $\#Unplanned$ by year to check if the reporting rate changes over time. In Column (4)-(6), the reporting rate is as high as 0.9 in 2015, decreases to 0.3 in 2016-2017, and further decreases to 0.2 in 2018-2019. This indicates an increasing underreporting rate over time, especially in and after 2016. Despite the limited time range of my sample, the findings emphasize the need for policy improvement.

Besides, I test whether the unreported abnormal outages are also polluting, or even more polluting than those reported events. The first test is to address the concern that it is sufficient only to reveal reported events. The second test is to check the strategic behavior of refinery plants to hide severe emissions. I replicate the first stage estimate in Table S11.2, and separate reported and unreported unplanned outages in Table S11.5. The first stage results hold only using plants in Texas, and the magnitude is more considerable. The unplanned shutdown day induces 0.7 units' increase in surrounding HCHO, equivalent to 8.2% of the mean. In Table S11.5, estimates on $UnplannedShut \times Unreported$ and those on $UnplannedShut \times Reported$ are positive. The latter one is statistically insignificant, probably due to the small proportion of reported events. Point estimates are larger for reported events. Surrounding pollution increases by 1 or 0.57 units when the unplanned outage is reported or not, 12.5% or 6.8% of the average pollution level. There is no strategic behavior to hide serious emissions and only report mild ones. While the reported outages are more serious shown in the large point estimates, we could not ignore those underreported events that also generate pollution spikes.

I also study the heterogeneous effects on downstream outcomes when the unplanned outage is reported or not. Results are summarized in Table S11.6 to S11.8. The proportion of pollution-related tweets increases when the unplanned outage is not reported. While this finding seems

counter-intuition, refinery plants submit excess emission forms within 24 hours of the pollutant release, and the information revealing happens further afterward. It is less likely that surrounding residents talk about the pollution displayed on the TCEQ website, but about the observed visual evidence or smell. Therefore, with similar limited information on pollutant release, the unreported events are more discussed or complained about by the public than the reported events. I find similar results on other outcomes in Table S11.6 Panel B and C, and Table S11.7 to S11.9. Offensive content and racist content on Twitter, crime events, medical expenditure, and hospital visits increase more when unplanned outages are not reported than reported ones. I conclude that self reporting in Texas is insufficient to capture abnormal pollutant releases from refineries, and unreported abnormal operations generate more adverse effects on surrounding residents.

In a similar vein, I test whether pollution impacts differ when unexpected outages trigger pollution alerts or not. The hypothesis is that behavioral responses are different when pollution alerts are on. People avoid going outside and adopt defensive measures, which mitigates adverse health effects. Compared with self-reported events that are organized and released weeks later, pollution alerts are informed to the public in advance,[98] so are more likely to induce behavioral responses. However, pollution alerts are based on the weather forecast and predicted high emission events rather than actual observed pollution signals (Mu et al., 2021), so they may not cover abnormal outages which are perfectly unexpected. At the same time, alert programs are implemented by state and local agencies and have incomplete coverage and unequal enforcement across the country. Air pollution alert data is from the EPA AirNow Action Day program.[99] I obtain a total of 32,532 alerts issued by 558 cities in 45 states 2014-2019. For each event, I am able to observe the event location recorded as coordinates, issue date, forecast date, forecasted AQI, and level of health concern. I overlay the actual alert area-days[100] and unexpected shutdown schedules to define whether an actual shutdown is warned or not. Among 101 sample plants, 40 of them experienced 13,835 events in their surrounding areas. However, only 11 out of 762 abnormal outage events are captured by pollution alerts, 1.44% of the total. This is consistent with the forward-looking pollution alert and the unexpectedness of abnormal operations. Table S11.10 shows the first stage results when abnormal outages are captured by pollution alerts or not. Estimates on $UnplannedShut \times Alerted$ are close to zero and statistically insignificant, probably due to a small number of events triggering alerts. Pollution increases are mainly observed in unalerted shutdowns. While local governments attempt to raise avoidance on unhealthy days,

[98] According to the AirNow Action Day data, 18,278 out of 32,532 alerts are scheduled one day in advance 2014-2019, 56.2% of the total. 4,112, 212, 87, and 6 are scheduled 2, 3, 4, and 5 days in advance respectively, 13.6% of the total. 9,839 (30.2%) are issued on the same day as actual high pollution days.

[99] Pollution alerts are implemented and recorded by each city or reporting area. Not all cities or local agencies join action day programs or have actually issued air quality alerts. Even though the pollution forecast is high, there may or may not be an associated alert. Cities that have alert programs could be found here: https://www.airnow.gov/aqi/action-days/

[100] I generate 20km buffer around alerted areas. If an abnormal refinery plant is located in the buffer, I consider the shutdown event triggered a pollution alert.

polluting events due to unexpected failure are not predicted, and the alert program is not sufficient to reduce exposure in this context.

S11.2 Recommended policy and cost-benefit analysis

Instead of self reporting, an alternative policy solution lies in utilization rates. I test the relationship between utilization rate and abnormal outages' frequency and severity. Refinery operation utilization data at the month-district level is obtained from EIA's Refinery Utilization and Capacity.[101] Table S11.4 shows a positive and insignificant correlation between plant utilization rate and the number of unplanned outage events. As the utilization rate increases by 10%, the number of unplanned outages increases by 0.1 for each plant in each month. Besides, I separately estimate equation (1) for each utilization group. Results in Table S11.3 show pollution spikes due to unplanned outages have a nonlinear relationship with utilization rate. Refineries with too low or too high utilization rates witness high pollution spikes, and high utilization groups generate more severe pollution. Figure S11.1 visualizes the U-shape relationship between utilization rate and estimated coefficients on $UnplannedShut$. Most refineries are currently operating on the right side of the optimal utilization rate with the lowest pollution increase. Therefore, I conclude that refinery plants should not operate at high utilization rates to reduce unplanned outage frequency and lower excess emissions in each unplanned outage. Current utilization rates may be refineries' private efficient conditions. In other words, at this utilization rate, their private costs of outage and revenue from oil products could generate the highest profits, assuming operators are rational. Given my findings, we should set an environmental regulation to reach social efficiency, i.e. lower utilization rates where private revenue equals private costs plus external costs. Refineries could either operate at lower utilization rates with existing plants or keep current production levels and raise capacities to lower the adverse impacts of unplanned outages.

Under a quadratic fit, the optimal utilization rate with the lowest pollution increase when unplanned outages is 89.7%. In reality, the average utilization rate of operating plants is 91.0%. Lowering utilization rates of those over-operating plants improve environmental conditions and generates benefits in downstream outcomes. However, it also reduces production and profits for refinery operators. To assess the net effect, I implement a cost-benefit analysis on the suggested policy of lowering utilization rates by 1.3%.[102]

If plants above the optimal rate reduce their utilization rates by 1.3%, 536 out of 864 plant-months have binding constraints. The number of unplanned outages in 2014-2019 would decrease from 468 to 436, 4.9% fewer than the actual level. Outage frequency aside, each event's induced

[101] Utilization data could be found here: https://www.eia.gov/dnav/pet/pet_pnp_unc_a_(na)_yup_pct_m.htm Plant-level utilization rate is not available. I use district-level utilization for further analysis.

[102] Lowering utilization rate is one recommended policy option, and there may be other alternatives to control unplanned outages like installing equipment to prevent catalyst release. I do not attempt to provide the best policy solution with readily available equipment, low cost for refinery operators, or little monitoring efforts for environmental regulators. Instead, I emphasize that it is feasible to control the outage problem.

air pollution increase is 0.32 units, 41.8% lower than the earlier level (0.55 units). The benefit of lowering utilization rates includes fewer crimes, lower medical expenditure, and fewer hospital visits.[103,104,105] Fewer events and lower pollution leads to 1.7 fewer crimes, $77.6 lower health expenditure, and 27.8 million fewer hospital visits in each county 2014-2019, 0.081%, 0.089% and 0.035% relative to the mean.[106] The equivalent monetary value of the latter two benefits is $0.04 million and $49 billion in the U.S. in each year.

The cost of lowering utilization rates lies in the lost production and profits. 536 treated plant-months are processing 149.5K barrels per day. With a reduced utilization rate by 1.3%, they would process 147.4K barrels per day instead. The equivalent lost revenue is $836 million in the U.S. in each year.[107]

Comparing the cost and benefit, the morbidity-related benefit of lowering the utilization rate by 1.3% exceeds the cost, $49 billion vs. $836 million. Despite a higher private cost on refineries, the externality of abnormal outage-induced pollution is avoided and social efficiency is improved. The benefit would be even higher if considering other non-morbidity benefits and spillover effects in domestic areas without refineries as well as in other countries. Therefore, I conclude that the social optimum with fewer abnormal outages is achievable, and we should set environmental regulations to correct the externality.[108]

[103]The benefit analysis only includes these three aspects with point estimates in Section 7 and 8. I do not attempt to provide a complete cost-benefit analysis, but would like to highlight that the restricted set of benefits already generate substantial monetary values.

[104]Regarding social media activities, to my knowledge, there is no estimate on the social cost of offensive tweets. Since there is no previous causal evidence that social media content affects real-world physical violence or mental health, no reference monetary value is available to include offensive tweets in this benefit analysis.

[105]In terms of crime, its social cost has a wide range of estimates varying across community characteristics and crime types. Property-related crimes are easy to evaluate but still depend on what is stolen, while person-related crimes depend primarily on the victims. Wickramasekera et al. (2015) reviews the estimated costs of crimes, and the total annual cost in the U.S. ranges from $450 billion to $3.2 trillion. Miller et al. (2021) shows in 2017, $2.6 trillion financial loss occurred due to 120 million crimes in the U.S.

[106]I calculate the effect as: effect = point estimates in reduced form result $\times$ 32 + point estimates in IV regression result $\times$ 0.23 $\times$ 436. The first term captures the effect of 32 avoided outages. In each unplanned outage, residents face 0.017 more crime events, $0.652/$0.068/$0.030 higher spending on medical products, and 2.488 more hospital visits, shown in Table 4, 7 and 8. The second part focuses on the remained 436 events. They generate 0.55 units' HCHO increase before and 0.32 units' increase with a lower utilization rate. Surrounding residents face 0.23 units' lower HCHO in the new 436 events. For each unit of HCHO increase, crime, medical expenditure, and hospital visits increase by 0.01, $0.318/$0.146/$0.070, and 1.977, shown in Table S7.13, S8.1 and S8.9.

[107]The lost oil production is calculated as: 2100 barrels per day $\times$ 365 days $\times$ 101 plants in the U.S. $\times$ 536/864 treated plant-months = 47.76 million barrels per year in the U.S. Refining 3 barrels of crude oil to produce and sell 2 barrels of gasoline and 1 barrel of diesel generates net profits averaging $17.50 per barrel of crude oil. Therefore, the lost profit is 47.76 million $\times$ $17.50 = $836 million.

[108]The cost-benefit analysis concludes that lowering utilization rates could generate higher benefits than costs and increase social welfare than the baseline scenario without regulating abnormal operations. Other potential policies may be more cost-effective than lowering utilization rates. My analysis does not attempt to compare policy options or identify which policy is the most cost-effective, but underscores it is not the second-best option not to regulate this issue.

Table S11.1: Self reporting of unplanned shutdowns in Texas (at the plant-month level)

	#Self-reported events					
	(1)	(2)	(3)	(4)	(5)	(6)
#Unplanned Outages	0.352***	0.348***	0.355***			
	(0.082)	(0.067)	(0.069)			
#Unplanned×Y2014				0.145	0.108	-0.094
				(0.240)	(0.187)	(0.245)
#Unplanned×Y2015				0.921***	0.931***	0.885***
				(0.317)	(0.272)	(0.256)
#Unplanned×Y2016				0.250	0.325**	0.339**
				(0.154)	(0.146)	(0.140)
#Unplanned×Y2017				.327*	.343**	.358**
				(.169)	(.155)	(.146)
#Unplanned×Y2018				.192	.11	.179
				(.191)	(.198)	(.236)
#Unplanned×Y2019				.222	.127	.271
				(.248)	(.257)	(.241)
Observations	1584	1584	1584	1584	1584	1584
R-square	0.011	0.234	0.255	0.017	0.241	0.261
Y-mean	0.923	0.923	0.923	0.923	0.923	0.923
Y-sd	1.168	1.168	1.168	1.168	1.168	1.168
Plant FEs		Y	Y		Y	Y
Year FEs			Y			Y
Month FEs			Y			Y

Notes: This table tests the underreport rate of refinery outages in Texas. Self-reported excess emission events are obtained from the Texas Commission on Environmental Quality (TCEQ). I calculate the number of self-reported events and actual abnormal outages at the plant-month level. The estimate in Column (1) means as the number of actual events increases by 1, reported events increase by 0.352, indicating the reporting system captures only 35.2% actual events. I add plant fixed effects in Column (2), and further add year and month fixed effects in Column (3). Column (4) to (6) separately estimate underreporting rate in each year. Standard errors are clustered at the plant level.

Table S11.2: First stage only using plants in Texas

	HCHO			
	(1)	(2)	(3)	(4)
PlannedShut	-0.012	-0.024	-0.015	0.002
	(0.358)	(0.463)	(0.351)	(0.359)
PlannedDowntime	-0.296***	-0.315*	-0.296***	-0.297***
	(0.087)	(0.172)	(0.087)	(0.076)
UnplannedShut	0.682***	0.738***	0.686***	0.681***
	(0.028)	(0.094)	(0.021)	(0.021)
UnplannedDowntime	-0.034	0.109	-0.034	-0.031
	(0.087)	(0.188)	(0.088)	(0.086)
OutageAnyPlant	-.362***	-.695***	-.361***	-.382***
	(.00498)	(.152)	(.00426)	(.00917)
Observations	23282	23282	23282	23282
R-square	0.124	0.136	0.124	0.125
Y-mean	8.316	8.316	8.316	8.316
Y-sd	5.486	5.486	5.486	5.486

Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S11.3: Heterogeneous effects across utilization rates

	HCHO				
	0-20th percentile	20-40th	40-60th	60-80th	80-100th
	(1)	(2)	(3)	(4)	(5)
UnplannedShut	1.196**	0.630**	-0.170	0.791***	1.338**
	(0.351)	(0.157)	(0.942)	(0.192)	(0.400)
Downtime	-0.188	-0.301**	-0.235*	-0.603***	-0.333
	(0.213)	(0.104)	(0.110)	(0.080)	(0.297)
OutageAnyPlant	0.162	-0.219	-0.068	-0.064	0.515
	(0.200)	(0.360)	(0.319)	(0.500)	(0.655)
Observations	30420	25772	10226	32179	6738
R-square	0.169	0.096	0.176	0.152	0.094
Y-mean	7.598	8.388	8.741	8.420	7.224
Y-sd	5.147	5.828	6.010	5.618	5.334
Plant FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
DOW FEs	Y	Y	Y	Y	Y
Mean utilization rate (%)	86.8	90.5	91.0	93.8	95.3

Notes: Standard errors are clustered at the plant level.

Table S11.4: Correlation of utilization and frequency of unplanned shutdowns (plant-month level)

	#Unplanned Shutdowns		
	(1)	(2)	(3)
Utilization	1.383	1.067	0.670
	(1.322)	(0.679)	(0.460)
Observations	864	864	864
R-square	0.004	0.231	0.249
Y-mean	0.763	0.763	0.763
Y-sd	1.731	1.731	1.731
Plant FEs		Y	Y
Year FEs			Y
Month FEs			Y

Notes: Utilization rate data is at the district-month level. I assign utilization rate to all plants in the same district in each month. Outcome variable is at the plant-month level, same as the key independent variable in Table S11.1. Standard errors are clustered at the district level.

Table S11.5: Heterogeneous effects on pollution when abnormal emissions reported or not

	HCHO			
	(1)	(2)	(3)	(4)
UnplannedShut × Unreported	0.570***	0.671***	0.574***	0.566***
	(0.213)	(0.050)	(0.202)	(0.200)
UnplannedShut × Reported	1.026	0.936	1.026	1.036
	(0.647)	(0.653)	(0.642)	(0.637)
Downtime	-0.296***	-0.313*	-0.296***	-0.297***
	(0.093)	(0.182)	(0.093)	(0.082)
OutageAnyPlant	-0.363***	-0.691***	-0.363***	-0.383**
	(0.002)	(0.143)	(0.001)	(0.006)
Observations	23282	23282	23282	23282
R-square	0.124	0.136	0.124	0.125
Y-mean	8.316	8.316	8.316	8.316
Y-sd	5.486	5.486	5.486	5.486
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S11.6: Heterogeneous effects on tweets when abnormal emissions reported or not

	Panel A: Proportion of tweets with air pollution keywords (in percentage)			
	(1)	(2)	(3)	(4)
UnplannedShut × Unreported	0.827**	0.942**	0.830**	0.830**
	(0.056)	(0.031)	(0.060)	(0.062)
UnplannedShut × Reported	-0.157	-0.011	-0.135	-0.136
	(0.145)	(0.080)	(0.174)	(0.162)
Downtime	-0.058	0.038	-0.057	-0.053
	(0.030)	(0.021)	(0.029)	(0.024)
OutageAnyPlant	-0.100	-0.108	-0.099	-0.098
	(0.146)	(0.210)	(0.147)	(0.135)
Observations	5891	5891	5891	5891
R-square	0.015	0.032	0.015	0.015
Y-mean	0.674	0.674	0.674	0.674
Y-sd	2.269	2.269	2.269	2.269
	Panel B: #Offensive tweets			
UnplannedShut × Unreported	56.419**	27.278***	56.344**	59.714**
	(28.290)	(8.633)	(28.680)	(26.278)
UnplannedShut × Reported	-8.358	56.841***	-7.019	-19.885
	(27.598)	(11.438)	(23.119)	(13.992)
Downtime	-38.175	-16.673	-38.180	-16.588
	(30.610)	(67.487)	(30.601)	(52.024)

OutageAnyPlant	3.232	-14.054	3.290	8.182
	(8.456)	(12.453)	(8.623)	(6.698)
Observations	3326	3326	3326	3326
R-square	0.466	0.509	0.466	0.474
Y-mean	147.700	147.700	147.700	147.700
Y-sd	412.766	412.766	412.766	412.766
	Panel C: #Racist tweets			
UnplannedShut × Unreported	10.433**	5.445***	10.426**	10.910**
	(4.671)	(1.638)	(4.726)	(4.395)
UnplannedShut × Reported	-4.134	6.722***	-4.015*	-5.861***
	(2.830)	(0.708)	(2.080)	(0.718)
Downtime	-5.271	-2.295	-5.272	-2.173
	(5.854)	(11.633)	(5.855)	(9.016)
OutageAnyPlant	0.621	-2.495	0.627	1.329
	(1.402)	(2.236)	(1.432)	(1.136)
Observations	3326	3326	3326	3326
R-square	0.424	0.466	0.424	0.430
Y-mean	22.178	22.178	22.178	22.178
Y-sd	70.701	70.701	70.701	70.701
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Table S11.7: Heterogeneous effects on hate crimes when abnormal emissions reported or not

	#Hate crime events against black people			
	(1)	(2)	(3)	(4)
UnplannedShut × Unreported	0.002	0.001**	0.002	0.002
	(0.001)	(0.000)	(0.001)	(0.001)
UnplannedShut × Reported	-0.009*	-0.008*	-0.009*	-0.009*
	(0.001)	(0.001)	(0.001)	(0.001)
Downtime	-0.003*	-0.003**	-0.003*	-0.003*
	(0.000)	(0.000)	(0.000)	(0.000)
OutageAnyPlant	0.001	-0.000	0.001*	0.000
	(0.000)	(0.001)	(0.000)	(0.000)
Observations	47762	47762	47762	47762
R-square	0.006	0.008	0.006	0.006
Y-mean	0.006	0.006	0.006	0.006
Y-sd	0.093	0.093	0.093	0.093
County FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the county level.

Table S11.8: Heterogeneous effects on expenditure when abnormal emissions reported or not

	Expenditure at the county-day level ($)		
	Cough remedies	Sinus remedies	Breathing aids
(Product module code)	(8425)	(8502)	(7790)
UnplannedShut × Unreported	0.334*	0.058	-0.005
	(0.033)	(0.035)	(0.003)
UnplannedShut × Reported	0.129	-0.217**	-0.014*
	(0.081)	(0.012)	(0.001)
Downtime	-0.261	-0.016**	-0.005
	(0.153)	(0.001)	(0.001)
OutageAnyPlant	0.259	0.009	0.011
	(0.131)	(0.006)	(0.004)
Observations	47762	47762	47762
R-square	0.173	0.036	0.009
Y-mean	1.751	0.205	0.032
Y-sd	6.036	1.441	0.543
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

Notes: Standard errors are clustered at the county level.

Table S11.9: Heterogeneous effects on foot traffic when abnormal emissions reported or not

	#Devices at the county-day level ($\times 10^{-5}$)	#Visits at the county-day level ($\times 10^{-5}$)	
		Amusement parks & recreational camps	General medical & surgical hospitals
	(1)	(2)	(3)
UnplannedShut × Unreported	0.045	-0.031	5.016**
	(0.044)	(0.013)	(0.179)
UnplannedShut × Reported	-0.063	-0.059*	1.521
	(0.083)	(0.005)	(1.229)
Downtime	0.590*	0.004	3.009*
	(0.078)	(0.025)	(0.374)
OutageAnyPlant	0.161	-0.014	1.361
	(0.067)	(0.007)	(0.831)
Observations	16016	16016	16016
R-square	0.958	0.739	0.850
Y-mean	3.497	0.372	22.860
Y-sd	4.839	0.457	39.886
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
Trends	Quadratic	Quadratic	Quadratic

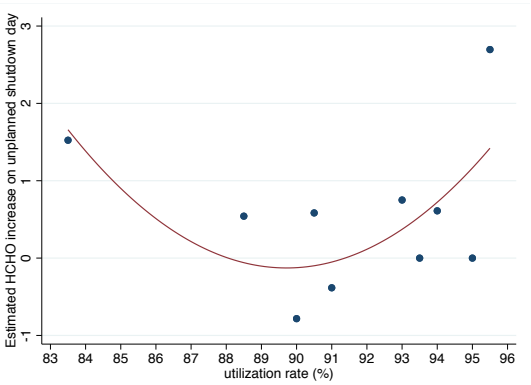
Notes: Standard errors are clustered at the county level.

Table S11.10: Heterogeneous effects on pollution when abnormal events trigger alerts or not

	HCHO			
	(1)	(2)	(3)	(4)
UnplannedShut × Alerted	0.009	0.263	-0.015	-0.009
	(0.840)	(0.746)	(0.870)	(0.866)
UnplannedShut × Unalerted	0.621**	0.625**	0.621**	0.621**
	(0.254)	(0.258)	(0.253)	(0.252)
Downtime	-0.352***	-0.373***	-0.354***	-0.355***
	(0.054)	(0.050)	(0.054)	(0.055)
OutageAnyPlant	0.032	0.002	0.013	0.001
	(0.105)	(0.171)	(0.104)	(0.108)
Observations	105335	105335	105335	105335
R-square	0.128	0.131	0.128	0.128
Y-mean	8.129	8.129	8.129	8.129
Y-sd	5.581	5.581	5.581	5.581
Plant FEs	Y	Y	Y	Y
Year FEs	Y		Y	Y
Month FEs	Y		Y	Y
Year-month FEs		Y		
DOW FEs	Y	Y	Y	Y
Trends			Linear	Quadratic

Notes: Standard errors are clustered at the plant level.

Figure S11.1: Heterogeneous effects across utilization rates



Notes: These figures display the estimated coefficients of *UnplannedShut* in equation (1), using plants in each 0.5% utilization rate interval. The navy scatters show raw values and the red plot is the quadratic fitted line.

S12 Permanent closure and reopening

S12.1 Data

This section quantifies the impact of recently retired and reopened refineries' normal operations on surrounding air quality. I use permanent closure and reopening schedules to answer this question. Data comes from EIA form 820 which documents the name, district, location, capacity, and month of the last operation of closed plants 1991-2020. For reopening events, I am able to observe months of the first operation of reopened plants. Locations are recorded as cities without precise coordinates.[109] Figure S12.1 displays counties with plants closure and reopening. 67 and 11 counties experience closure and reopening events, and 5 counties experience both. They are mainly located in the south and the west U.S.

S12.2 Empirical strategy

I use an event study design to test the impact of plants' permanent closure and reopening on air pollution at the county-month level. When plants are permanently closed, I expect pollution levels to decrease; when plants are permanently reactivated, I expect pollution to increase. The dynamic event study model is specified in the following equation:

$$P_{it} = \sum_{k=-24}^{24} \alpha_k D_{it}^k + Time_t + County_i + \varepsilon_{it} \tag{10}$$

where the sample includes air pollution at the county-month level. The time range is 24 months before to 24 months after the last operation of the closed plants; for reactivation, the time range is 24 months before to 24 months after the first operation. Counties include those with at least one closed or reopened plant. Outcome variable P_{it} refers to the average HCHO level in county i in month t. D_{it}^k is a dummy that equals one if month t is the kth month relative to the last operation month of the closed plant in county i and zero otherwise. The dummy for k=-1 is omitted. $Time_t$ includes year, month fixed effects, and quadratic time trends. They are to capture unobserved nationwide conditions that correlate with time. I also add county fixed effects on the right-hand side to control for county-level static characteristics. To account for within-county across-time autocorrelation in the error term, I cluster standard errors at the county level.[110]

Coefficients α_1,...,α_{24} capture the impact of permanent closure on surrounding pollution relative to baseline pollution in month -1 when the plant is operating. Those for restart events indicate the impact of permanent reopening on surrounding pollution compared with that in month -1 when the plant is not operating. The identifying assumption is the exogeneity of plants' permanent closure and reopening. In other words, no other confounders that may affect

[109] Given the inexact location, I use county-level air quality for further analysis on permanent closure and reopening events. Besides, I restrict the analysis using events between January 2005 and December 2019, given the time range of pollution data.

[110] For later analysis on temporary normal and abnormal outages, I cluster standard errors at the plant level in the first stage and reduced form analysis when studying tweet responses. I cluster standard errors at the county level with crimes and health outcomes at dependent variables. The level is consistent with the location unit of outcome variables.

surrounding pollution exist at the same time with permanent events. This assumption is not testable given the unobserved counterfactual. Instead, I test whether pre-trends before treatments are significantly different from zero. My identifying assumption holds if the estimated α_{-24},...,α_{-2} are statistically insignificant.

S12.3 Results

I assess the contribution of recently retired and reopened plants to local air pollution. Results from estimating equation (10) are reported in Table S12.1. Estimates on *Post* in Column (1)-(3) capture the effects of permanent closure on surrounding HCHO. They are not statistically different from zero with county, year and month fixed effects controlled. This suggests the negligible effects of plant retirement on air quality. In Column (4)-(6), I find similar imprecise estimates on *Post*, indicating the little contribution of reopened plants to surrounding pollution.

Figure S12.2 shows the event study figures. I report estimated coefficients and their 95% confidence intervals in each month from two years before to two years after the permanent events. The large standard errors in the pre-periods in both panels confirm the exogeneity of permanent closure and reopening events. In the post-period, the estimated coefficients range from -2 to 2 units but are not statistically different from zero. The figures indicate that the pollution patterns are the same as that in month -1.

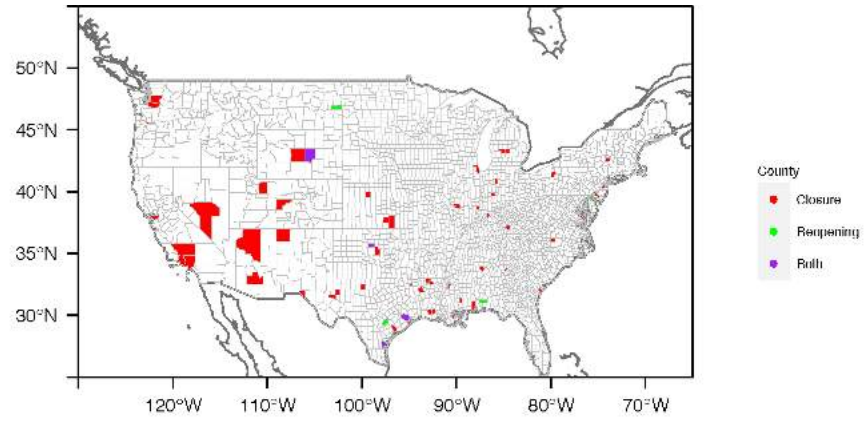
Table S12.2 reports estimated results for the dynamic model. I report the average treatment effects for each three-month bin and drop month -3 to 0. Consistent with Figure S12.2, all the estimates in Column (1), (3), (4) and (6) remain statistically insignificant. To sum up, I conclude that neither permanent closure nor reopening generates changes in county-level HCHO.

I also use monitor data to test the impact of permanent events on other criteria pollutants. Results in Table S12.3 show similar negligible effects on SO_2, $PM_{2.5}$, benzene, NO_2 and O_3. In all these panels, estimates on *Post* in the restarting events are insignificant or not robust. For closure events, only benzene shows a decrease of 0.31 parts per billion carbon in the post-period, 19.8% relative to the main. The other pollutants are not improved by the plant retirement. Note that the estimation has a smaller sample size than that in Table S12.1, due to the insufficient coverage of ambient air quality monitors near refinery plants shown in Figure S3.2.

One potential reason for the negligible effect of plant retirement is the low operating level in the last operation months. Production shifts to other refinery plants to ensure the market supply of oil products, and plants ramp down until they officially get closed. Similarly, reopened plants may operate at low utilization rates in the first operating months. These could result in the minor impact of permanent events on surrounding air pollution. The other potential reason is the pollutant leak even after the operation stops.[111] After the operation ends, the unprocessed crude materials or catalysts including hydrocarbons are not cleaned up. The retirement of refineries generates no benefit, and nearby residents still bear the environmental cost of the refinery.

[111]The leakage problem after plant closure is documented in news articles. For example, 'the Philadelphia Energy Solutions refinery continued to emit concentrations of benzene far above EPA's limit for the carcinogenic gas even half a year after a June 2019 catastrophic fire shut down the 150-year-old oil processing facility'. Source: E&E News, March 2020.

Figure S12.1: Counties with permanently closed and reopened plants



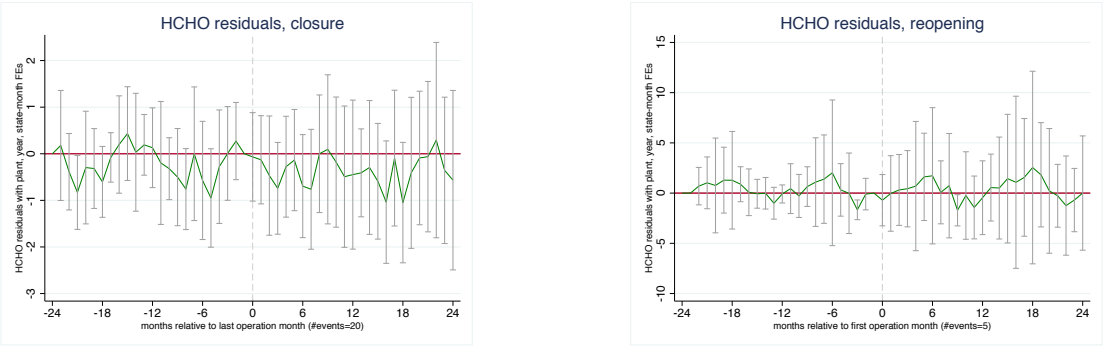
Notes: This figure shows 67 counties in red with permanently closed plants, 11 counties in green with reopened plants, and 5 counties in purple with both types of permanent events.

Table S12.1: Effects of permanent closure and reopening on surrounding HCHO

	HCHO					
	Closure			Reopening		
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.093	-0.244**	-0.092	-0.241	0.997	-0.230
	(0.122)	(0.104)	(0.123)	(0.501)	(0.716)	(0.497)
Observations	965	965	965	244	244	244
R-square	0.594	0.799	0.594	0.573	0.713	0.573
Y-mean	7.785	7.785	7.785	8.369	8.369	8.369
Y-sd	2.678	2.678	2.678	2.491	2.491	2.491
County FEs	Y		Y	Y		Y
Year FEs	Y	Y	Y	Y	Y	Y
Month FEs	Y		Y	Y		Y
County-Month FEs		Y			Y	
Quadratic trends			Y			Y

Notes: Standard errors are clustered at the county level.

Figure S12.2: Effects of permanent closure and reopening on surrounding HCHO



(a) Closure events

(b) Reopening events

Notes: These figures display the treatment effect of refinery plants' permanent closure and reopening on surrounding HCHO levels in each month relative to the pre-treatment period (month -1). I only use balanced events without any missing satellite reports in the 49-month event window. The green lines display the estimated coefficients after controlling for plant, year, state-month fixed effects, and the gray bars show 95% confidence intervals.

Table S12.2: Effects of permanent closure and reopening on HCHO, dynamic model

	HCHO					
	Closure			Reopening		
	(1)	(2)	(3)	(4)	(5)	(6)
Time [-24,-21)	0.427	0.084	0.432	0.531	-0.906**	0.518
	(0.263)	(0.267)	(0.262)	(0.540)	(0.258)	(0.530)
[-21,-18)	0.658	-0.309	0.627	-0.154	-0.364	-0.166
	(0.576)	(0.282)	(0.576)	(1.689)	(0.802)	(1.710)
[-18,-15)	0.145	0.027	0.124	-0.678	-0.621	-0.681
	(0.433)	(0.340)	(0.445)	(1.025)	(0.697)	(1.048)
[-15,-12)	0.070	0.173	0.067	0.372	-0.677	0.395
	(0.240)	(0.259)	(0.240)	(0.434)	(0.525)	(0.455)
[-12,-9)	.394	.00198	.402	.104	-.254	.104
	(.305)	(.224)	(.31)	(.972)	(.644)	(.938)
[-9,-6)	.738	-.278	.717	-.309	.46	-.31
	(.51)	(.329)	(.511)	(1.53)	(.763)	(1.53)
[-6,-3)	-.319	-.441	-.336	-.739	.283	-.747
	(.417)	(.381)	(.426)	(1.4)	(.881)	(1.43)
(0,3]	.418	-.297	.42	-.387	.617	-.357
	(.421)	(.222)	(.422)	(1.18)	(.943)	(1.13)
(3,6]	.656	-.277	.642	-.74	1.35	-.731
	(.538)	(.219)	(.542)	(1.75)	(1.11)	(1.75)
(6,9]	-.397	-.266	-.411	-.789	.694	-.795
	(.344)	(.346)	(.353)	(.875)	(.711)	(.89)
(9,12]	-.31	-.456	-.314	-.257	.654	-.261
	(.308)	(.328)	(.309)	(.847)	(1.1)	(.851)
(12,15]	.433	-.321*	.44	-.012	1.87	.0308
	(.343)	(.178)	(.349)	(.497)	(.881)	(.437)
(15,18]	.293	-.673**	.283	-.565	2.48	-.545
	(.541)	(.239)	(.539)	(1.34)	(1.38)	(1.32)
(18,21]	-.369	-.272	-.391	-.0065	2.43*	-.0195
	(.374)	(.325)	(.394)	(.878)	(.871)	(.894)
(21,24]	-.0814	-.327	-.102	-.236	1.64	-.267
	(.276)	(.296)	(.27)	(.539)	(.972)	(.573)
Observations	965	965	965	244	244	244
R-square	0.608	0.802	0.608	0.588	0.742	0.588
Y-mean	7.785	7.785	7.785	8.369	8.369	8.369
Y-sd	2.678	2.678	2.678	2.491	2.491	2.491
County FEs	Y		Y	Y		Y
Year FEs	Y	Y	Y	Y	Y	Y
Month FEs	Y		Y	Y		Y
County-Month FEs		Y			Y	
Quadratic trends			Y			Y

Notes: Standard errors are clustered at the county level.

Table S12.3: Effects of permanent closure and reopening on other pollutants using monitor data

	Closure			Reopening		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: SO_2						
Post	-0.303	-0.205	-0.250	0.421	0.228	0.471
	(0.285)	(0.151)	(0.298)	(0.575)	(0.451)	(0.596)
Observations	367	367	367	234	234	234
R-square	0.507	0.663	0.520	0.458	0.543	0.476
Y-mean	1.405	1.405	1.405	0.800	0.800	0.800
Y-sd	1.344	1.344	1.344	1.217	1.217	1.217
Panel B: $PM_{2.5}$						
Post	0.203	0.592	0.203	0.338	0.172	0.368
	(0.247)	(0.620)	(0.245)	(0.997)	(0.478)	(1.025)
Observations	621	621	621	244	244	244
R-square	0.606	0.616	0.606	0.443	0.658	0.444
Y-mean	9.436	9.436	9.436	10.314	10.314	10.314
Y-sd	5.042	5.042	5.042	2.768	2.768	2.768
Panel C: Benzene						
Post	-0.313*	-0.336**	-0.311*	0.050	0.429***	0.046
	(0.144)	(0.121)	(0.142)	(0.096)	(0.023)	(0.093)
Observations	386	386	386	135	135	135
R-square	0.484	0.575	0.485	0.672	0.674	0.674
Y-mean	1.562	1.562	1.562	1.780	1.780	1.780
Y-sd	0.956	0.956	0.956	0.722	0.722	0.722
Panel D: NO_2						
Post	0.006	-0.777	0.011	-0.673	-0.990	-0.596
	(0.281)	(1.061)	(0.269)	(2.172)	(2.362)	(2.236)
Observations	392	392	392	143	143	143
R-square	0.844	0.707	0.844	0.806	0.838	0.808
Y-mean	10.448	10.448	10.448	10.875	10.875	10.875
Y-sd	5.062	5.062	5.062	3.744	3.744	3.744
Panel E: O_3						
Post	0.000	0.001	0.000	-0.001	-0.001	-0.001
	(0.000)	(0.001)	(0.000)	(0.002)	(0.001)	(0.002)
Observations	593	593	593	241	241	241
R-square	0.680	0.825	0.680	0.419	0.814	0.423
Y-mean	0.031	0.031	0.031	0.028	0.028	0.028
Y-sd	0.009	0.009	0.009	0.007	0.007	0.007
County FEs	Y		Y	Y		Y
Year FEs	Y	Y	Y	Y	Y	Y
Month FEs	Y		Y	Y		Y
County-Month FEs		Y			Y	
Quadratic trends			Y			Y

Notes: Standard errors are clustered at the county level.

S13 Keywords and example tweets

Air pollution keywords (all in lowercase):

- smog, air pollution, air quality, clean air, haze, particulate matter, particle, particulates, emission, pm25, pm2.5, pm10, so2, no2, aqi, o3, ozone, visibility, voc, volatile organic compound, benzene, halogenated compounds, esters, aromatic hydrocarbons
- #airpollution, #airquality, #cleanair, #particulatematter, #volatileorganiccompound, #halogenatedcompounds, #aromatichydrocarbons
- odor pollution, odor nuisance, odorous, odorant, noxious odor, noxious smell, unpleasant odor, unpleasant smell, chemical odor, chemical smell
- #odorpollution, #odornuisance, #noxiousodor, #noxioussmell, #unpleasantodor, #unpleasantsmell, #chemicalodor, #chemicalsmell

Health keywords:

- cough, asthma, stroke, heart attack, heart ailment, lung cancer, lung disease, breast cancer, liver cancer, pancreatic cancer
- #heartattack, #heartailment, #lungcancer, #lungdisease, #breastcancer, #livercancer, #pancreaticcancer

Tweets with air pollution keywords (UTC time):

- 'What's up with this air pollution in the Gulfgate area? #stinky #Pasadena' created at 2014-11-15 18:17:43
- 'morning haze. #htx #htxphotographer #shotoniphone #shotoniphonex #editedwithvsco' 2018-04-04 12:29:37
- 'That's a lot of smoke in the sky. #airpollution #petrochemicalfire #airquality #sigh #houston' 2019-03-19 00:30:45
- 'Weirdly smoggy day, but on the water nonetheless. #saltlife #saltwater #water #southtexas #southtx #sotx #texas @Corpus Christi, Texas' 2019-02-17 01:37:27

Tweets with health keywords:

- 'I think I'm developing asthma' 2014-03-30 06:07:02
- 'I just had a stroke' 2014-09-12 19:02:14
- 'Shout out to myself for almost having an asthma attack today at soccer tryouts. not ok for the first day' 2014-12-01 22:37:17

Offensive tweets:

- 'Stuck at Walmart waiting on a manager to complain about the stupid B*TCH who checked us out.' 2015-01-01 00:31:05
- 'Not only are you a liar, your rude as well.' 2017-12-20 10:04:19

Racist tweets:

- 'That's some dark ass sh*t.' 2018-10-30 04:00:42
- 'Done with relationships, done with communication from guys, done with talking to you stupid ass n*, done catching feelings for you n* & lastly done feeling sorry for you n*.' 2019-04-03 18:31:45