Title
The Regulation and Health Effects of Air Pollution

Author
Mullins, Jamie Thomas

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Regulation and Health Effects of Air Pollution

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Jamie Thomas Mullins

Committee in charge:

Professor Joshua Graff Zivin, Chair
Professor Prashant Bharadwaj
Professor Richard Carson
Professor Gordon Dahl
Professor Craig McIntosh
Professor Gordon McCord

2015

The dissertation of Jamie Thomas Mullins is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

To my family and friends for supporting me during this process. In particular, to my wonderful, beautiful, patient, supportive, and loving wife for moving to San Diego and encouraging me every step along the way. Lexie, I love you forever. Also, to our beautiful daughter, Leena, without whom the preparations of this disseration would not have been quite so memorable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

A special thanks to my dissertation committee members: Josh Graff Zivin, Prashant Bharadwaj, Gordon Dahl, Craig McIntosh, Gordon McCord, and Richard Carson. Also a huge thanks to the attendees in the UCSD Environmental Lunch Seminar and the International/Development Seminar. Additionally I would like to thank Matthew Gibson, Matthew Neidell, Marc Jacobsen, Ted Groves, and Karthik Muralidharan for their comments on this research. Finally, thanks to the first year crew that got me through qualifying exams: Andy Brownback, Sarojini Hirshleifer, Arman Rezaee, and Matt Gibson.

Chapter 1, in full, is currently under review for publication. Jamie Mullins, the dissertation author, was the primary investigator and author of this paper.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Jamie Mullins, the dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is a reprint of the material as it appears in the American Journal of Agricultural Economics, September 2014, aau081. Mullins, Jamie and Bharadwaj, Prashant, 2014. The dissertation author was the primary investigator and author of this paper.

VITA

2006	B.A. in Industrial Engineering and Operations Research, University of California, Berkeley
2012	M.A. in Economics, University of California, San Diego
2015	Ph.D. in Economics, University of California, San Diego

PUBLICATIONS

Mullins, Jamie (with Prashant Bharadwaj) "Effects of Short-Term Measures to Curb Air Pollution: Evidence from Santiago, Chile." American Journal of Agricultural Economics (2014): aau081.

Mullins, Jamie (with Janet Currie, Joshua Graff Zivin, and Matthew J. Neidell) "What Do We Know About Short-and Long-Term Effects of Early-Life Exposure to Pollution?." Annual Review of Resource Economics 6.1 (2014): 217-247.

Mullins, Jamie (with Joshua Graff Zivin). "Vessel buybacks in fisheries: The role of auction and financing structures." Marine Policy 53 (2015): 188-197

ABSTRACT OF THE DISSERTATION

The Regulation and Health Effects of Air Pollution

by

Jamie Thomas Mullins

Doctor of Philosophy in Economics

University of California, San Diego, 2015

Professor Joshua Graff Zivin, Chair

This dissertation focuses on the relationship between air pollution and human activity. The first chapter examines whether firms are better incentivized to clean up their atmospheric emissions via a regulatory scheme based on an absolute or relative standard. The second chapter examines intercollegiate track & field outcomes to identify an adaptation effect of training in higher ambient ozone levels. The third chapter looks at the efficacy of short term government interventions in Santiago, Chile, which sought to reduce the severity and health effects of spikes in urban air pollution through a series of government protocols enacted to address poor air quality.

Chapter 1

Motivating Emissions Cleanup: Absolute vs. Relative Performance Standards

1.1 Introduction

The comparative merits of assessments based on Relative versus Absolute performance standards have long been debated.[1] For example, will grading on a curve or an absolute scale motivate students to study harder, or will workplace compensation based on a fixed performance threshold be better or worse for productivity compared to a tournament scheme where the returns to extra effort are uncertain? Such questions are relevant in many other contexts as well, including a wide range of policy settings. For instance, the Clean Air Act relies on a series of Absolute thresholds (one for each of six criteria pollutants) to assign counties to non-attainment status, while the Race to the Top initiative incentivized states to adopt K-12 educational reforms by awarding funds through a competition between states.[2] The central analysis of this paper focuses on a natural policy experiment in which a Relative standard and an Absolute standard were both used (in sequence) to support the same emissions cleanup program. Comparing outcomes of the otherwise unchanged program under the two regimes provides an opportunity to examine and compare the efficacies of the approaches in motivating effort.

This comparison is of particular interest, as the common application of Absolute and Relative performance standards in closely related scenarios highlights both the substitutability of the approaches and the lack of consensus over which best motivates regulated agents. While a number of papers have sought to compare the motivational properties of Absolute and Relative stan-

[1]Relative standards - under which incentives are assigned by comparing individual performance against the performance of peers - are also known as norm-referenced standards, while Absolute standards - under which incentives are assigned by comparing individual performance against fixed threshold levels of performance - are sometimes referred to as criterion-referenced standards.

[2]Non-attainment status under the Clean Air Act involves a higher level of regulatory scrutiny and the implementation of a plan to achieve compliance with the violated National Ambient Air Quality Standard (or risk losing some forms of federal funding). The Race to the Top initiative awarded points to states for the adoption of specified types of reform, and the states earning the most points during each of three rounds were awarded shares of $4.35 billion in funds.

dards theoretically,[3] very few have empirically compared outcomes between the standards (Czibor et al., 2014 and Paredes, 2012 being notable exceptions). Empirical comparisons of Absolute and Relative performance standards are rare for three main reasons. First, the behavior incentivized by such standards is often not directly observable, and proxies - when available - tend to introduce additional factors to the analysis which may hamper causal analysis.[4] Second, agent heterogeneity complicates causal identification for cases in which the performance of each agent is not observed under both standards and random assignment is not possible. In such settings inference must rely on either a cross group comparison of the same task or a within agent comparison of different tasks, both of which introduce potentially confounding factors to the analysis. Finally, any comparison of the Relative and Absolute approaches - including those relying on random assignment of the performance standards - must ensure that the strictness of the compared standards is equal, such that observed differences in behavior are attributable to the standards rather than differential regime stringency.

This paper addresses each of these potential issues through the examination of repeated applications of a regulatory program aimed at motivating the cleanup of stationary sources of air pollution in Chile. Specifically, the program was intended to motivate reductions in the concentration of particulate matter (hereafter "concentration") in the atmospheric emissions of stationary sources of air pollution (hereafter "plants") in the Santiago Metropolitan Region. Toward this goal, the emissions cleanup program first relied on a Relative standard (punishing plants responsible for 30% of emissions) and subsequently an Absolute standard (based on a fixed concentration threshold). Under both

[3]Most notably Lazear and Rosen (1981), Green and Stokey (1983), and Nalebuff and Stiglitz (1983) examine threshold-based and competitive wage structures, while Becker and Rosen (1992) and Dubey and Geanakoplos (2010) compare Absolute versus curved gradings schemes in the classroom.

[4]For instance, both Czibor et al. (2014) and Paredes (2012) rely on student test scores as a measure of effort, but the link between test scores and effort is clearly modulated by a number of unobservable factors, including student ability, that are difficult to properly account for.

regimes, the concentrations of plant emissions have been closely monitored, and thus the incentivized behavior is directly observable in the program's administrative data. Furthermore, this emissions cleanup program has been effectively in force for nearly two decades with each year serving as a separate performance standard implementation, and with the change from a Relative to an Absolute standard occurring in 2001. Thus, for each of nearly 4,000 plants, we observe the year-on-year concentration reductions (hereafter "abatement") under both performance standards, allowing for the effective control of unobservable plant-specific factors that might contribute to behavioral responses. Finally, when the Chilean program transitioned from the Relative standard in 2000 to the Absolute standard in 2001, the fixed threshold under the Absolute standard was set to match the effective concentration threshold for punishment in 2000, and the regulated population remained largely unchanged. Together these facts provide an ideal natural experiment in which uncertainty surrounding the level of the threshold under the Relative standard is resolved while the stringency of the applied standard is held fixed.

My analysis begins with a conceptual model based on the idea that it is the uncertainty surrounding the threshold level under the Relative regime that differentiates it from the Absolute standard. My main finding is that agents react to this difference in regulatory uncertainty in one of two distinct ways according to their ex ante identifiable type. Low compliance-cost types increase effort in response to the greater uncertainty of the Relative standard, while those with high compliance-costs under-supply effort in the face of uncertainty, leading this type to put forth more effort under the Absolute regime than under the Relative. As a result, the aggregate comparison of emissions cleanup efforts between the regimes is dependent on the densities of these two agent types in the population of interest.

My empirical analysis compares plant abatement under the last year of the Relative standard (2000) to abatement in the first year of the Absolute standard (2001), exploiting the constant punishment threshold across these two years. I find that aggregate abatement was 6.4 percentage points lower (on a basis of of around 19% per year) under the Absolute regime than under the Relative, and that this result is driven by a 14.1 percentage point average abatement reduction among plants identified ex ante as having low compliance-costs. In contrast, I find that plants with high compliance-costs undertake approximately 27.7 percentage points more abatement under the Absolute regime than under the Relative. The large share of plants identified as being low compliance-cost types (approximately 80 percent of the sample) drives the sign of the aggregate result as anticipated.

The central examination of this paper treats all agents (i.e, plants) as equivalent decision makers, which is the key to comparing the efficacies of our two standards in generating effort. However, all concentration reductions do not contribute equally to the improvement of air quality, because total mass of particulate emissions is the product of concentration and emissions volume. I also therefore undertake a policy assessment of the effect of the 2001 change in performance standards on the annual reductions of total particulate matter emissions. This assessment is particularly important given the difficulty of regulating atmospheric emissions in the context of an emerging economy (see for instance: Greenstone et al., 2011 and Pande et al., 2012), and I find that the 2001 change from a Relative performance standard to an Absolute one reduced the effectiveness of the emissions cleanup program by 14.5 percentage points (the equivalent of a 41% reduction).

Finally, I provide an illustrative example of how changes in the distribution of compliance-costs affect the aggregate comparison of effort under the Absolute and Relative performance standards. This exercise demonstrates that an increase in the share of high compliance-cost types (akin to an increase in regulatory strictness) leads to a rise in the comparative efficacy of the Absolute regime versus the Relative standard.[5] This example highlights

[5] Others have examined the effects of stricter standards within either an Absolute or Relative regime, but none have made the comparison between the approaches. See for example: Kang (1985), Betts and Grogger (2003), Figlio and Lucas (2004), and Babcock

the importance of taking into account the cost of regulatory compliance for the agents of interest. The higher the average costs, the better an Absolute regime is likely to perform, but the efficacy of a Relative standard will comparatively improve as the average costs of compliance fall amongst the regulated. Taken together, my results have significant implications for a wide range of regulators,[6] and more generally for any principal seeking to incentivize a group of heterogeneous agents.

The remainder of the paper is structured as follows. In Section 1.2, I discuss prior literature which addresses uncertainty and environmental regulation. Section 1.3 describes the Chilean emissions cleanup program and the data used in my empirical investigation. I derive the conceptual model and its implications in Section 1.4, and describe my empirical methods in Section 1.5. Section 2.6 presents the main results of the empirical analysis, and I present a number of robustness checks in Section 1.7. Section 1.8 provides an illustrative example of the impact of a change in the distribution of agent types on the comparison of Absolute and Relative standards. Section 1.9 concludes.

1.2 Environmental Regulation and Uncertainty

This paper focuses on uncertainty in the level of the punishment threshold as the defining difference between Absolute (no threshold uncertainty) and Relative (with threshold uncertainty) performance standards. Both Bandyopadhyay and Horowitz (2006) and Shimshack and Ward (2008) focus on a different possible avenue of uncertainty in emissions regulation, examining plant behavior in the face of an uncertain link between cleanup effort and achieved abatement. Similar to my aggregate empirical result, both papers find that plants put forth more effort to comply with regulations when un-

(2010).

[6]For instance, all 50 states would have to develop and implement individual carbon emissions regulations for energy generation facilities under the newly proposed Clean Power Plan (EPA, 2014).

certainty is increased. Regulators too must deal with uncertainty in policy settings, and it is different degrees of regulator uncertainty regarding costs and benefits of compliance that motivate the Weitzman (1974) comparison of price vs. quantity regulatory mechanisms. Although, Lazear (2006) does not directly address environmental regulation, my findings closely mirror his theoretical conclusions that agent types with low and high costs of compliance are differently motivated by stochastic incentives, with well-defined requirements (akin to Absolute standards) driving more effort from those with high costs of compliance.

There are also broader implications of the presence or absence of uncertainty in regulation. For instance, Relative performance standards save regulators from having to set threshold performance levels which are both challenging and attainable because the mechanism sets the standard level (Becker and Rosen, 1992). Additionally, the threshold uncertainty of the Relative regime may preempt regulatory corruption of the types identified by Duflo et al. (2013) in India and Ghanem and Zhang (2014) in China. The uncertainty of the threshold level under a Relative regime makes such gaming much more difficult, but it also requires a full census of the regulated population each time incentives are to be assigned (a significant and costly task in any developing regulatory framework). Alternatively, an Absolute regime allows for as-monitored assignment of punishments or rewards in real time,[7] a feature which may be especially valuable given that enforcement of environmental regulations is notoriously difficult to begin with (Heyes, 2000). Additionally, the lower level of regulatory uncertainty faced by plants under an Absolute regime may serve to spur investment (Yang et al., 2004). As evidenced by the widespread application of the Absolute and Relative standards in closely re-

[7]The idea that a more effective incentive mechanism may have higher monitoring costs is akin to the reasoning in Shavell (2009) as to why strict liability rules are not more common despite their greater ability to induce efficient behavior modification (compared to negligence standards). In a unilateral accident, determination of fault under strict liability requires full regulator knowledge of all prior relevant behaviors, while such determinations made under a negligence standard require only knowledge of actions at the time of incident.

lated contexts, there are many reasons why one might be preferred in a given setting. The goal of this paper is to shed light on one important aspect of the comparison between these two approaches, namely how much compliance effort is driven by each, and under what circumstances one approach might lead to more effort than the other.

1.3 The Emissions Cleanup Program and Administrative Data

Beginning in the early 1990s, the Chilean government undertook a number of significant steps to address the poor air quality in the Santiago Metropolitan Region. Between 1990-1994, a series of rules were implemented that together require stationary industrial and institutional emitters of vapor, smoke, or steam (referred to in this paper as "plants") to annually obtain and report measurements of particulate matter concentrations in their atmospheric emissions while operating at full capacity.[8] Such measurements were required to be undertaken each year by government-approved third-party auditors following methods laid out by the EPA of the United States (Environmental Protection Agency, Method 5), and reported to the Chilean government by a pre-specified date early each year (see Appendix Table 1.11 for an annual time line). Between measurement cycles, a group within the Metropolitan Region's Environmental Health Service undertook audits of plant emissions in order to identify plants operating at concentrations different from those reported. Although violating plants have been found (and punished) through the years, the reported emissions characteristics are generally considered by regulators

[8]These rules include: mandated annual measurement of emissions concentrations by all plants (Supreme Decree No. 32/1990, hereafter SD32/1990); an obligation for plants to comply with a concentration standard (SD4/92); and an annual requirement for plants to demonstrate (through measurement reporting) compliance with the standard of SD4/92 (Resolution 15027/1994 of the Metropolitan Region Department of Health). All these rules, are concerned with particulate matter of ten micrometers in diameter or less, commonly referred to as PM_{10}.

and outside researchers to be highly reliable (Palacios and Chávez, 2005).

In order to incentivize plants to reduce the concentrations of their emissions, each year a group of the dirtiest plants (i.e., those with the highest reported concentration levels) are assigned to a Shutdown List, which requires that they cease emitting operations on days of particularly poor air quality in the upcoming pollution season (winter in Santiago). Importantly for this examination, the performance standard used to assign plants to the Shutdown List was Relative until 2000 and has been Absolute ever since.

Under the Relative regime, regulators followed a heuristic which assigned the plants with the highest operational concentrations (regardless of plant size) to the Shutdown List until the plants on the Shutdown List were responsible for 30% of aggregate Total Mass Emissions ("TME")[9] of particulate matter in the Metropolitan Region. This standard is akin to grading on a curve with a 70% pass rate, except that instead of all participants having equal weights, larger plants count for more in the tally of failed (or passing) grades. The concentration that I refer to as the "implied" or "effective" threshold under the Relative standard is the lowest concentration of a plant on the Shutdown List. The concentration of the cleanest punished plant serves as an effective punishment threshold because all plants with higher concentrations are punished and all those with lower concentrations are not. This implied threshold fell each year under the Relative standard.

Since the 2001 change in performance standard, the government has assigned plants to the Shutdown List based on an Absolute threshold of $32mg/m^3$, which was set to match the implied punishment threshold in the last year of the Relative performance standard, and has not changed since.[10] Un-

[9]TME is a secondary measure of particulate matter emissions that is calculated as the product of a plant's emissions concentration and "flow". While flow is a measure of the volume of emissions output (per hour), the Chilean rules require that it be a measure of emissions capacity rather than actual operational levels. Thus, TME is a measure of the capacity to emit particulate matter rather than a measure of actual emissions.

[10]The differential effects of a year-on-year ratcheting in the threshold level versus a fixed level are controlled for in the empirical investigation and discussed further in Section 1.5.

der the Absolute regime, plants reporting concentrations above this threshold are placed on the Shutdown List, while those below it are not.

Importantly, the punishment regime represented by assignment to the Shutdown List did not change with the performance standard in 2001, and in fact has not changed since the emissions cleanup program began. Assignment to the Shutdown List for a given year means that a plant is required to cease all emitting operations on days in that year which are designated by the Chilean government as Environmental Episodes (hereafter "Episodes"). Such Episodes serve as the official identification of days on which air quality in Santiago is expected (or in some cases unexpectedly realized) to be particularity poor.

Episodes come in three levels of severity: Alert, Pre-Emergency, and Emergency. Alert level Episodes do not require any plant shutdowns and are thus omitted from further discussion and analysis in this paper. Although Emergency Episodes are included in the analysis of this paper, they are exceedingly rare (only two Emergency Episodes occurred during the study period), and I have therefore simply grouped them together with Pre-Emergency Episodes (of which there are 103 instances in the study period). Going forward I will use the term Episode to refer primarily to Pre-Emergency level Episodes, though the two Emergency Episodes are also included in this grouping for the analysis.[11]

While an Episode announcement precipitates a broad range of measures, such announcements are germane to this paper only in that they trigger 24-hours of mandated cessation of all emitting processes by plants on the Shutdown List.[12] It is worth noting that while the exact number of Episodes in any given year (and thus the cost of punishment in that year) is stochastic

[11] Although Emergency Episodes do invoke higher shutdown requirements, I will set this aside due to the infrequency of observed Emergency Episodes, and treat Emergency Episodes as equivalent to Pre-Emergency Episodes in the analysis.

[12] Protocols and restrictions differed by the level of Episode announced, but all announcements are completely independent of the Shutdown List. See Appendix Table 1.12 for more details on the different Episode levels. For more information on Episodes in general, and an empirical examination of their effectiveness in reducing air pollution and mortality in the short term, please see Mullins and Bharadwaj (2014).

and unknown ex-ante, during the study period of this paper (1995-2010), an average of 6.5 Episodes were announced per year.[13]

The data used as the basis for the empirical examination of this paper are the comprehensive annual reports of concentration, flow (i.e., emissions volume), and TME levels for each plant in the Santiago Metropolitan Region from 1995-2010. These data were maintained by a Chilean government entity within the Ministry of the Environment known as the Program for Fixed Source Emissions Control (hereafter "PROCEFF") which was responsible for regulating plants. Because the emissions cleanup program was not fully in place until 1994, the first year in which data are available is 1995, and 1996 is the first year in which the data are comprehensive (i.e., include all active plants). Information on plant owner, address, comuna (i.e., neighborhood), emitting process type, fuel type, and fuel usage are also included in the data. Four types of emissions process are analyzed in this paper: Heaters (used for heating air and water), Industrial Boilers, Baking Ovens, and Other Emitting Processes. Electricity generators were regulated separately until 2004 and are thus excluded from the analysis. The data provided on each plant allows for the linking of observations across years so that each specific plant can be followed through the study period. Table 2.5 contains annual summary statistics of the data used in the empirical analysis of this paper.

Generally, the aggregate annual statistics show marked reductions in emissions concentrations and emissions capacity (measured as TME), with only mild reductions in the average volume of plant emissions, and a significant growth in the number of plants in the Santiago Metropolitan Region. The growth in the number of plants - from 3,671 in 1996 to 8,804 in 2010 - reflects the general growth in the Chilean economy - per capita GDP rose from $5,168 in 1996 to $12,682 in 2010 (measured in 2014 USD - World Bank, World Development Indicators). Given the difficulty of balancing environmen-

[13]Plant Abatement does not significantly predict Episode counts suggesting that plants are not able to anticipate or effectively impact the number of Episodes ultimately announced (results not presented).

tal considerations with economic development (see for instance: Greenstone et al., 2011 and Pande et al., 2012), it is worth noting the apparent success of the Chilean emissions program in cleaning up the average plant (mean concentrations fell) and reducing the total emissions of particulate matter (TME shrank), even as the number of plants more than doubled.

1.3.1 Policy Context

In 1998, the Relative standard was amended so that plants responsible for 30% of TME were assigned to the Shutdown List. Previously the share had been only 20%, so this change amounted to an increase in strictness. As reported in Table 2.5, this increase in regulatory strictness appears to have been quickly internalized into plant concentrations as the mean annual abatement level jumped nearly 10-fold from 1998 to 1999, the number of plants undertaking positive abatement (hereafter "Abaters") nearly tripled, and the number of active plants fell (for the only time in the data). Such immediate reactions suggest that plants were aware of, and reactive to, the incentives of the emissions cleanup program. Further, such rapid and striking behavioral changes suggest that the effects of the 1999 policy change were internalized well before the change in performance standards examined by this paper, and thus can be reasonably omitted from my examination of plant behavior in 2000 versus 2001.

In addition to the policies which are the focus of this investigation, three other programs may have influenced concentration choices of individual plants during the study period. The first of these - which will be referred to as the Clean Fuel Program ("CFP") - assigned fixed concentration levels to certain small (i.e., $flow < 1000m^3/hr$) plants using fuels from an approved list, which exempted them from the annual requirement for a full emissions measurement (though continued compliance with the CFP requirements had to be demonstrated). The assigned concentrations were engineering estimates of actual emissions concentrations, and both the estimates and the list of approved fuels were updated periodically. I include CFP plants in the analysis even though their emissions concentrations were not measured every year, because their assigned concentrations serve as a reasonable proxy (at least on average) for their actual concentrations.

Secondly, the government applied harsh punishments to plants that violated maximum concentration ceilings ($112mg/m^3$ for large plants and $56mg/m^3$ for small plants), which may have affected concentration setting decisions among plants with particularly high operational concentrations. By the time of the change in standards, we see few such plants, likely due to the enforcement of these concentration ceilings. While I don't expect such ceilings to affect my central analysis, the possibility will be addressed via controls in my empirical specifications (for further discussion see: Section 1.5).

The last program that warrants special attention is the 1992 implementation of a cap-and-trade program on the total particulate matter emissions of large (i.e., $flow \geq 1000m^3/hr$) plants. This program garnered much interest among economists and environmental policy makers, but is widely regarded to have been ineffective at motivating plant emissions reductions or cleanup because the cap was set too high and enforcement ranged from lax to nonexistent (Montero et al., 2002; Coria, 2009; Palacios and Chávez, 2005). As a result, I will not further consider this program in the analysis that follows.

1.4 Conceptual Model

With the empirical setting laid out, I now develop a conceptual model describing optimal plant concentration setting decisions under Relative and Absolute performance standards. Through the comparison of optimal concentration setting under the two regimes, the model generates predictions for the response to the change from a Relative performance standard to an Absolute standard that took place in Chile. The model's prediction of two distinct groups of responses to the switch from Relative to Absolute performance standards provides a basis for the investigation of type-specific heterogeneous effects.

Each period, t, is modeled as having two stages. In the first stage, each plant, i, selects an operational concentration, ϕ_i^t, and in the second stage punishments are assigned and administered. Plants select an operational concentration in each period by maximizing expected profits for the period, solving:

$$\max_{\phi_i^t} E\left[\pi_i^t(\phi_i^t)\right] \tag{1.1}$$

Decomposing the profit function, $\pi_i^t(\phi)$, and omitting the expectation notation yields:

$$\max_{\phi_i^t} \left[Rev_i^t - OC_i^t(\phi_i^t) - PC_i^t(\phi_i^t)\right] \tag{1.2}$$

where Rev_i^t is annual revenue, $OC_i^t(\phi)$ is operational costs for a given operational concentration, and $PC_i^t(\phi)$ is the anticipated cost of punishment, which also depends on the chosen operational concentration.

In the absence of regulation or incentives to cleanup emissions (i.e., when $PC_i^t(\phi) = 0 \ \forall \ \phi$), each plant would operate at some operational-cost minimizing concentration intrinsic to its emitting process, production technology, and fuel. Such emissions technologies are assumed fixed across the two years examined by the model, and I label the resulting characteristic concentration as the plant's "baseline concentration", denoted as ρ_i.[14] Noting that $\arg\max_\phi \left[Rev_i^t - OC_i^t(\phi)\right] = \rho_i$, I define "deviation costs" to be the increase in operational costs faced by a plant operating at a concentration other than its baseline, formally: $DC_i(\phi) = OC_i^t(\phi) - OC_i^t(\rho_i)$. Deviation costs therefore account for revenue losses attributable to non-baseline operational concentra-

[14]The assumption of fixed technology across the time of the regime change is based on the short time-frame at issue and the empirical evidence suggesting that adoption of new technologies was neither a cause, nor significant channel, for the changes in compliance behaviors that are the focus of this investigation. For further discussion, see Section 1.7.

tions so that annual revenue, Rev_i^t, should be thought of as annual revenue from operating at baseline concentration, and therefore independent of ϕ_i^t. Now I rewrite the plant's optimization problem as:

$$\begin{gathered}\max_{\phi_i^t}\left[Rev_i^t - OC_i^t(\rho_i) - DC_i(\phi_i^t) - PC_i^t(\phi_i^t)\right] \\ or \\ \min_{\phi_i^t}\left[DC_i(\phi_i^t) + PC_i^t(\phi_i^t)\right]\end{gathered} \tag{1.3}$$

Note, $\phi_i^t > \rho_i$ is never optimal because any operational concentration above the baseline will incur strictly higher deviation costs and a weakly higher probability of punishment compared to $\phi_i^t = \rho_i$. Thus, $DC_i(\phi_i^t)$ is strictly decreasing in ϕ_i^t over the feasible range of optimal concentrations, since $DC_i(\phi_i^t)$ is assumed to be strictly increasing in the distance from ϕ_i^t to ρ_i. The anticipated costs of punishment, $PC_i^t(\phi_i^t)$, are strictly increasing in ϕ_i^t on the interval of possible levels of the effective threshold, defined as $\left[\underline{\phi_i^t}, \overline{\phi_i^t}\right]$, and flat outside of this interval.

1.4.1 Optimal Concentration Setting

For any considered operational concentration, the cost of punishment is unknown as long as the punishment threshold is unknown. Plants are therefore modeled as making concentration setting decisions based on expected punishment costs, relying on a probability distribution over possible levels of the punishment threshold. Punishment costs, $PC_i^t(\phi)$, can thus be decomposed to: $F_i^t(\phi)R_i^t(\phi)N_i^t$, where $F_i^t(\phi)$ is the cumulative distribution function (and $f_i^t(\phi)$ the probability density function) for the level of the regulatory threshold, $R_i^t(\phi)$ is daily revenue at a given operational concentration,[15] and N_i^t is the expected fraction of days on which Episodes will be announced (and

[15] $R_i^t(\phi)$ equals $\left[Rev_i^t - OC_i^t(\phi_i^t)\right]/365$. Although R_i^t depends on ϕ_i^t, this dependence will not be specified going forward as the outcomes of the model do not materially hinge on this relationship.

thus shutdowns required for those on the Shutdown List). The uncertainty surrounding the threshold under the Relative regime arises from each plant's imperfect information regarding the costs faced by other plants, and I assume that possible thresholds anticipated by plant i are bounded on $\left[\underline{\phi_i^t}, \overline{\phi_i^t}\right]$, where $0 \leq \underline{\phi_i^t} \leq \overline{\phi_i^t} < \infty$. Additionally, I assume that the distribution of possible threshold levels is continuous, and may be plant specific. The scenario in which the true distribution of the threshold is known (and therefore common across plants) is thus captured as a sub-case.[16]

From the latter representation in Equation 1.4, I decompose $PC_i^t(\phi_i^t)$, differentiate, and set the result equal to zero. The resulting first order condition can be broken up over the intervals of optimal concentration, ϕ_i^{t*}, as follows:

$$DC_i^{'}(\phi_i^{t*}) = 0 \quad if \quad \phi_i^{t*} < \underline{\phi_i^t} \tag{1.4}$$

$$DC_i^{'}(\phi_i^{t*}) + f_i^t(\phi_i^{t*})R_i^t N_i^t = 0 \quad if \quad \underline{\phi_i^t} \leq \phi_i^{t*} \leq \overline{\phi_i^t} \tag{1.5}$$

$$DC_i^{'}(\phi_i^{t*}) = 0 \quad if \quad \overline{\phi_i^t} < \phi_i^{t*} \tag{1.6}$$

Thus we see that the optimal concentration should be set such that the marginal cost of achieving a lower concentration level, $DC_i^{'}(\phi)$, exactly equals the value of the decrease in punishment probability from further concentration reductions, $f_i^t(\phi_i^{t*})R_i^t N_i^t$. Under the Absolute regime the threshold level, ϕ_{TH}^t, is known at the time of concentration setting, thus: 1.) $\underline{\phi_i^t} = \overline{\phi_i^t} = \phi_{TH}^t$ and 2.) $f_i^t(\phi) = 0 \; \forall \; \phi \neq \phi_{TH}^t$ and $f_i^t(\phi_{TH}^t) = 1$. Solving for the optimal concentration under the Relative and Absolute regimes (denoted by $t = R$ and $t = A$ respectively) yields the following optimal value functions (OVFs) in terms of baseline concentrations:[17]

[16] Alternatively, the Relative regime could be modeled as an all-pay, multi-prize contest with head starts. All analyses of such contests (See for example: Clark and Riis, n.d.; Siegel, 2009; and Siegel, 2011) rely on strong assumptions of perfect information which are particularly unrealistic between thousands of plants that vary in size, process type, and ownership structure. Given the size and diversity of the population of plants, I do not directly address strategic interaction in the analysis that follows. Nevertheless, strategic considerations could be driving the definition of $F_i^t(\phi)$, and thus resulting behaviors would

Equation 7: Relative OVF *Equation 8: Absolute OVF*

$$\phi_i^{R*} = \begin{cases} \rho_i & if\ \rho_i < \underline{\phi_i^R} \\ \\ \rho_i - \Lambda_i & if\ \underline{\phi_i^R} + \Lambda_i \leq \rho_i \leq \overline{\phi_i^R} + \Lambda_i \\ \\ \rho_i & if\ \rho_i > \overline{\phi_i^R} + \Lambda_i \end{cases} \qquad \phi_i^{A*} = \begin{cases} \rho_i & if\ \rho_i < \phi_{TH}^A \\ \\ \phi_{TH}^A & if\ \phi_{TH}^A \leq \rho_i \leq \phi_{TH}^A + reach_i^A \\ \\ \rho_i & if\ \rho_i > \phi_{TH}^A + reach_i^A \end{cases}$$

where $\Lambda_i\left(\phi_i^{t*}, R_i^t N_i^t\right) = f_i^t(\phi_i^{t*}) R_i^t N_i^t$, and can be thought of as the additional abatement achievable for the value of a marginal reduction in anticipated punishment costs at the optimal concentration level.[18] Note that Λ_i may depend on ϕ_i^{t*}, and that $\Lambda_i \geq 0$. "Reach" is defined as the distance-from-baseline achievable for the cost of certain punishment, in the current context: $reach_i^t(R_i^t N_i^t) = R_i^t N_i^t$.[19]

In addition to fulfilling the equimarginal principle laid out above, both OVFs intuitively imply that a plant with a very high or very low ρ_i, should simply operate at its baseline concentration. Further qualitative characterizations of the OVFs in Equations 7 and 8 are as follows:

- Under the Relative regime:
  - a plant with a baseline concentration (ρ_i) below the lower bound of

still be well described by this model.

[17]For clarity of exposition, I impose the functional form assumption: $DC_i(\phi_i^t) = (\phi_i^t - \rho_i^t)^2/2$. This precise functional form is not necessary for the results that follow, but makes their representation straight forward.

[18]Equation 7 assumes a continuous belief distribution that falls smoothly to zero mass at both ends of its support. If there is positive mass at or near the lower end of a plant's belief support (as in a uniform distribution), there is a gap in the interval covered by lines 1 and 2 of Equation 7 because $\underline{\phi_i^R} \neq \underline{\phi_i^R} + \Lambda_i$. In such cases, the additional line: $\phi_i^{R*} = \underline{\phi_i^R}\ \ if\ \ \underline{\phi_i^R} \leq \rho_i^T < \underline{\phi_i^R} + \Lambda$, must be added to Equation 7.

[19]Both the forms of $\Lambda_i\left(\phi_i^{t*}, R_i^t N_i^t\right)$ and $reach_i^t(R_i^t N_i^t)$ are simplified by the functional form chosen for deviation costs, but both terms enter the OVFs similarly even under more general assumptions. Reach can also be thought of as the maximum amount plant i could abate for the value of certain avoidance of punishment. My definition of $reach_i^t$ is related, but not identical to that of Siegel (2009).

the threshold probability interval ($\underline{\phi_i^R}$) should operate at its baseline. Operation at any other concentration would incur positive cost and offer no expected benefit.

- a plant with baseline concentration (ρ_i) in or somewhat above the range in which the threshold might fall ($>$ $\underline{\phi_i^R}$ and $\leq$ $\overline{\phi_i^R} + \Lambda_i$), should operate at a concentration below its baseline. Such a plant should heuristically continue to consider lower concentration levels until the marginal cost of achieving a lower level exactly equals the value of the perceived decrease in punishment probability from further concentration reductions (formally: $\phi_i^{t*} = \phi_i^t \; s.t. \; DC_i'(\phi_i^t) = -f_i^t(\phi_i^t)R_i^tN_i^t$).
- a plant with baseline concentration (ρ_i) well above its support interval for the threshold distribution ($\rho_i > \overline{\phi_i^R} + \Lambda_i$) should also operate at its baseline concentration as the benefit from any lower concentration - as measured via the reduction in expected probability of punishment - is outweighed by the costs.

- Under the Absolute regime:
  - a plant with a baseline concentration(ρ_i) below the (known) threshold (ϕ_{TH}^A) should operate at baseline concentration.
  - a plant with a baseline concentration (ρ_i) above the threshold (ϕ_{TH}^A), but for whom the cost of operating at the threshold is less than the cost of punishment (formally: $DC_i(\phi_{TH}^A) < R_i^tN_i^t$), should operate at the threshold concentration. We say that the threshold is "within reach" for such plants.
  - a plant with a high baseline concentration ($\rho_i > \phi_{TH}^A + reach_i^A$) should operate at its baseline concentration, as the cost of avoiding punishment - by operating at the threshold - is greater than the

expected costs of certain punishment (such plants are said to be "out of reach" of the threshold).

The ultimate goal of characterizing optimal concentration setting behavior is to compare a plant's behaviors under the two performance standards. In order for this comparison to be meaningful and empirically relevant, I let the punishment threshold set under the Absolute regime equal that realized under the Relative regime. Panel A of Figure 1.1 plots modeled operational concentrations (on the vertical axis) against baseline concentrations (on the horizontal axis) under an Absolute (dotted line) and a Relative (solid line) regime when the distribution of the threshold under the Relative regime is $Beta(3, 2.5)$ on the interval $\left[\underline{\phi_i^t}, \overline{\phi_i^t}\right]$.[20] The distance between the OVFs characterizes the difference in plant behavior predicted under the two regimes. From any initial concentration level, the difference in abatement between the two regimes will be characterized by the difference between the OVFs at the relevant baseline concentration, and thus we can examine the changes in behavior anticipated by the model for a change from a Relative standard to an Absolute by examining the distance from the solid to the dotted line on the graph. Panel B of Figure 1.1 plots the difference between the two OVFs by baseline concentration, and thus represents the predicted change in abatement expected in response to a change in performance standards from Relative to Absolute.

In both panels of Figure 1.1, four distinct groupings of differential plant behavior are readily apparent. Intuitively, plants with very low baseline concentration (i.e., below $\underline{\phi_i^t}$) have no motivation to abate under either regime, and thus the difference in abatement between the regimes is zero. The same is true for plants with high baseline concentrations (i.e., above

[20] In cases when positive weight is placed on the upper endpoint of the interval, as under a Uniform distribution, I additionally assume that $\phi_{TH}^A + reach_i^A \geq \overline{\phi_i^R} + \Lambda_i$. This assumption is met if the plant believes the likelihood of the threshold falling at, or very near the upper bound of the interval is small, and the support interval itself is not too wide. Alternatively, this assumption is met if the cost of certain punishment is fairly large. As both of these conditions are likely to be satisfied in the empirical setting of interest, I do not include discussion of outcomes under the alternative assumption.

$\phi^t_{TH} + reach^t_i$). The interesting behavioral differences exist for plants that are somewhat near the threshold. For plants on the interval $\left[\underline{\phi^t_i}, \phi^t_{TH} + \Lambda_i\right]$, the Absolute OVF is above the Relative OVF implying that abatement incentives are stronger under the Relative regime (as optimal concentrations are lower). The converse is true when a plant's baseline concentration falls on the interval: $(\phi^t_{TH} + \Lambda_i, \phi^t_{TH} + reach^t_i]$. Over this interval the Relative OVF is above the Absolute OVF, yielding optimal concentrations which are higher - and motivation to undertake abatement which is lower - under the Absolute regime than the Relative. To summarize, a plant-specific critical value is deteriorated as the baseline concentration at which the OVFs cross. A plant with baseline concentration below this critical value - which occurs at: $\phi^t_{TH} + \Lambda_i$ - are expected to abate (weakly) less in response to a change from a Relative performance standard to an Absolute, while it is anticipated that plants with a baseline concentration above this critical value will abate (weakly) more after the change in performance standards that occurred in the Chilean context.

1.4.2 Mapping Model to Empirics

Although baseline concentrations are not observable in the data, plants with baseline concentrations above versus below their individual critical values of $\phi^t_{TH}+\Lambda_i$ can be differentiated based on concentration setting behavior under the Relative regime (i.e., ex ante to the policy change of interest). Specifically, if a plant operated with a concentration that ultimately fell below the implied threshold of the Relative regime, then the plant has a baseline concentration less than or equal to its specific value of: $\phi^t_{TH} + \Lambda_i$, which is the crossing point of its two OVFs. I label such plants as "Bin A" plants. Similarly, plants with $\rho_i > \phi^t_{TH} + \Lambda_i$ can be ϕ^R_i such that it ends up being $> \phi^t_{TH}$ once the threshold is revealed. I label such plants as "Bin B" plants. Thus, the model predicts that Bin A plants will weakly reduce abatement efforts in response to the change in performance standards while Bin B plants will weakly increase abatement efforts. Note that both the modeled criteria ($\rho_i \lesseqgtr \phi^t_{TH} + \Lambda_i$) and the empirical mapping ($\phi^R_i \lesseqgtr \phi^t_{TH}$) rely on the realized level of the threshold, and that this mapping is possible because the baseline concentration at which the OVFs cross is always the same level at which the Relative OVF crosses the fixed threshold level. Because the Bin A and B groups separate plants based on the sign of the effect of interest, the bins provide a solid basis for the investigation of heterogeneous plant responses to they policy change of interest.[21] See Panel A of Figure 1.1 for a graphical illustration of the OVF cross point and the empirical mappings of Bins A and B.

For illustrative purposes, I have only varied the level of baseline concentration in the discussion and illustrations thus far, holding constant the distributions of possible thresholds under the Relative regime, deviation costs, and costs associated with assignment to the Shutdown List. Empirically however, the levels of all three of these factors likely vary simultaneously with baseline concentration between plants.[22] The conceptual model was developed with this in mind, and the bin division criteria, $\rho_i \lesseqgtr \phi^t_{TH} + \Lambda_i$, already takes these additional dimensions of variability into account through the plant specific term, Λ_i.[23] As a result, the empirical mapping described above already correctly sorts plants into bins based on predicted behavior, but I will change the way I reference the resulting plant types going forward. Rather than describing Bins A and B simply as containing plants with low and high baseline concentrations respectively (and thereby glossing over the other dimensions

[21] Also, there are no clear empirical identifiers for plants with very high ($\rho_i > \phi^t_{TH} + reach^t_i$) or very low ($\rho_i < \underline{\phi^t_i}$) baseline concentrations. Thus no further separation of plant types can be made based on predictions from the model.

[22] All graphs and comparisons up to this point can be thought of as representing a group of plants that share these characteristics (captured in Λ_i, $reach^t_i$, and the OVFs), and differ only in their baseline concentration.

[23] The assumed functional form of deviation costs provides this clean representation of the bin division criteria. More generally it would include the term $DC'^{-1}_i\left(f(\phi^{t*}_i)R^t_iN^t_i\right)$, which is the inverse of the derivative of the deviation costs function. What is important is not the clean form of the term, but the fact that it takes into account plant-specific baseline concentration, threshold distribution, deviation costs, and costs associated with assignment to the Shutdown List.

of possible variation between plants which contribute to bin assignment), the labels of low and high compliance-costs will be used instead. Plants in Bin A are labeled as having low compliance-costs, while those in Bin B are labeled as high compliance-cost plants.[24] Note that Bin A plants are said to have low compliance-costs because they have either: 1.) a low baseline concentration (compared to the realized threshold level) or 2.) baseline concentration, costs, and belief structures (along with the realized threshold level) such that the threshold was achieved even under the uncertainty of the Relative regime. Conversely, the high compliance-costs of Bin B plants are a manifestation of the combination of their baseline concentrations, costs and belief structures (along with the realized threshold level) which did not allow these plants to cost effectively achieve the ultimately realized threshold under the uncertainty of the Relative regime.

According to this conceptual model, the answer to which standard leads to more abatement differs for those plants that are low versus high compliance-cost types. The comparison of total abatement between the two regimes will thus be determined by the distribution of plants between these two types (i.e., across compliance-costs) in the regulated population. Because the punishment rate was so low under the Chilean Relative regime - only about 20% of plants[25] were on the Shutdown List in 2000 - the large majority of plants are identified as low compliance-cost types (i.e., Bin A members) at the time of the regime change. With such a high share of low compliance-cost types, we expect that the aggregate effect of the 2001 change from a Relative to an Absolute regime

[24] Alternatively, we can think of low and high compliance-cost plants as having a quantity: $\rho_i - \Lambda_i$ that is below or above (respectively) the fixed threshold under the two performance standards being compared. If plants have a value of this quantity below the threshold, they will have selected an operational concentration under the Relative regime that ends up being below the revealed threshold, and will be placed in Bin A, identifying them as having low compliance-costs. Conversely, plants with $\rho_i - \Lambda_i > \phi_{TH}$ will be labeled as high compliance-cost types.

[25] Although plants responsible for 30% of TME of particulate matter were punished each year, differences in plant emissions volumes meant that a smaller share of the plant count were on the list in the last year of the Relative regime.

will be driven by the effect on the low compliance-cost type, a thus we expect to see a reduction in abatement on average.

1.5 Empirical Approach

The empirical examination focuses on the 2000 to 2001 regulatory change from a Relative to an Absolute performance standard. The first step in this process is a straightforward comparison of compliance behaviors in 2000 to those in 2001, both graphically and statistically. Other than the regime change, these two years were identical in terms of regulation, and thus any difference in abatement can be thought of as motivated by differences between regimes. Aggregate and type-specific comparisons of abatement in 2000 to abatement in 2001 are specified simply as:

$$\mathcal{Y}_{it} = \alpha + \beta * APS_t + \epsilon_{it} \tag{1.9}$$

$$\mathcal{Y}_{it} = \mathbf{type_i'}\hat{\alpha} + APS_t * \mathbf{type_i'}\hat{\beta} + \varepsilon_{it} \tag{1.10}$$

These regressions are estimated with abatement as the outcome variable, $\mathcal{Y}_{it}$, and with a sample that includes only 2000 and 2001 data. Abatement is measured as the year-on-year change in logged concentration levels for a given plant. APS_t is an indicator for the Absolute performance standard (or equivalently for $year > 2000$), and $\mathbf{type_i}$ is a vector of two indicators identifying each plant as either a low or high compliance-cost type at the time of the regime change (according to the empirical mapping described in the previous section). α is a constant and $\hat{\alpha}$ is a vector of type-specific constants. The plant/year specific error terms are ϵ_{it} and ε_{it}.

In the first specification, the coefficient of interest is on the APS_t variable and can be interpreted as the percentage point change in abatement (which is itself a percent change) associated with the change in performance standards from Relative to Absolute. The coefficients in the vector $\hat{\beta}$ in the second specification capture the percentage point change in abatement for each type separately, and allow for the examination of heterogeneous effects of the regime change. Finally, it is worth noting that my use of a first-differenced outcome variable removes all time-invariant plant-specific factors from the analysis, precluding the usefulness of plant fixed effects.

While the unit of observation in the data is plant/year, it is common for several plants to be owned by a single entity. Concentration decisions made at two different plants by a single owner cannot be thought of as independent observations of behavior, and thus all results are reported with standard errors clustered by owner (referred to hereafter as a "firm").

1.5.1 Main Specification

While the regressions above are appealing for their simplicity, a number of factors which might act on abatement decisions at the time of the regime change are not addressed. First and foremost among these factors is the automatic ratcheting effect that occurred under the Relative regime, through which the implied punishment threshold fell each year. As a result, abatement decisions under the Relative regime were made partially due to the expectation of a lower threshold in the coming year compared to the prior year, and can be expected to exhibit a trend which should be accounted for in the pre-2001 period. In order to control for time trends in the data, a longer time frame, from 1995-2010, will be analyzed. On such a broader time horizon, emissions technologies were becoming cleaner and cleaning approaches were becoming cheaper. These dynamics suggest that there might be some non-zero trend in abatement even if no regulation was in force to encourage it. Ignoring such a trend in an empirical estimation risks inappropriately attributing concentration reductions due to exogenous secular forces to the regime change.

To address the ratcheting threshold before the regime change and the secular cleanup trend over the 1995-2010 study period, I undertake additional regression analyses which take into account separate quadratic time trends in the pre and post-regime-change periods. When type-specific analysis is done, such time trends are included for each type.[26] A collection of plant-specific controls are also included to address changes in plant characteristics that might impact concentration setting behavior, but are unrelated to the policy change of interest.

$$\begin{aligned}\mathcal{Y}_{it} &= \alpha + \beta * APS_t + \mathbf{X}'_{\mathbf{it}}\hat{\gamma} + \lambda_1 * t_t + \lambda_2 * t_t^2 + \lambda_3 * APS_t * t_t \\ &\quad + \lambda_4 * APS_t * t_t^2 + \varepsilon_{it} \end{aligned} \tag{1.11}$$

$$\begin{aligned}\mathcal{Y}_{it} &= \mathbf{type}'_{\mathbf{i}}\hat{\alpha} + APS_t * \mathbf{type}'_{\mathbf{i}}\hat{\beta} + \mathbf{X}'_{\mathbf{it}}\hat{\gamma} + t_t * \mathbf{type}'_{\mathbf{i}}\widehat{\lambda}_1 + t_t^2 * \mathbf{type}'_{\mathbf{i}}\widehat{\lambda}_2 \\ &\quad + APS_t * t_t * \mathbf{type}'_{\mathbf{i}}\widehat{\lambda}_3 + APS_t * t_t^2 * \mathbf{type}'_{\mathbf{i}}\widehat{\lambda}_4 + \epsilon_{it} \end{aligned} \tag{1.12}$$

Just as in the earlier specifications, APS_t and $\mathbf{type_i}$ are, respectively, an indicator for the Absolute performance standard and a vector of indicator functions for plant type of either low or high compliance-costs. The term $\mathbf{X_{it}}$ is a vector of plant/year specific "Additional Controls" that include start-of-period flow and fuel consumption levels, indicators for whether switching to natural gas was an option and whether a plant participated in any sort of Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Time trends are addressed with t_t as a year counter with $year = 2000$ set to $t_t = 0$. ε_{it} and ϵ_{it} are again plant/year specific error terms. These specifications are evaluated on the full sample of plants running from 1995-2010 in order to best identify the pre and post regime-change time trends.

In addition to abatement, I also estimate the regressions specified in Equations 1.11 and 1.12 as Probit models with the outcome variables being

[26]Quadratic time trends are used because of the trends observed in the raw data (see Figures 1.2) and due to the relatively short length of the pre and post periods. Nevertheless, the use of alternative trend specifications does not change the main thrust of the results and conclusions (results not presented).

indicators for plant i being an Abater (i.e., plant operating at a lower concentration in year t than in year $t-1$) or De-Abater (i.e., plant operating at a higher concentration in the current year than in the preceding year) in year t. These estimations serve as another means of examining how the different performance standards affected abatement behaviors.

1.5.2 Weighted Regressions

In order to assess the impact of the change in performance standards on air quality, I will weight each plant observation by its daily emissions volume and re-estimate the main regressions.[27] This approach will serve as a policy assessment for the 2001 change in performance standards by assessing the effects of the regime change on the total amount of particulate matter emitted by stationary sources in the Santiago Metropolitan Region.

1.6 Results

Figure 1.2 presents three graphs of the mean abatement by year under the Relative regime and under the Absolute regime. The largest panel shows the smoothed evolution of abatement increasing considerably year-on-year through 2000, and then dropping from above 20% per year to less than 15% with the regime change in 2001. The obvious trends and strong slope change at the time of the regime-change highlight the importance of controlling for time trends in the pre and post periods, but even setting aside the change in slopes and acknowledging that other factors need to be controlled for, the main panel of Figure 1.2 suggests a negative aggregate impact of the

[27]Daily emissions volume is estimated as the product of flow and daily hours of operation. Because flow reports a measurement of maximum (or potential) emissions volume, I am using reported capacities as a proxy for actual emissions volumes. If all plants operate at capacity, or at a common share of capacity, then the analysis will be accurately weighted. For the case in which we can only say that larger capacity plants tend to emit higher volumes (as seems likely), my analysis will still give a good flavor for the comparative effects of performance standards on reductions in the total mass of emitted particulate matter.

regime change on abatement, while the two smaller panels show sharply heterogeneous effects by compliance-cost type.

The three panels of Figure 1.2 make it clear that the aggregate effect of the change in performance standards closely mirrors that among low compliance-cost plants. This fits well with our understanding of the relative densities of the two types of plants in the regulated population. In 1999 and 2000, the plants responsible for 30 percent of TME were placed on the Shutdown List. This resulted in about 20 percent of plants (by count, not TME) being assigned to the Shutdown List in 2000, meaning that approximately 80 percent of plants had a 2000 concentration below the 2000 threshold and are thus identified as low compliance-cost types (see the first column of Table 1.2 for plant counts). As the model anticipates, it is clear from the second panel of Figure 1.2 that low compliance-cost plants undertake less abatement after the regime change, and since such plants represent around 80 percent of all plants, the aggregate plant behavioral response to the regime change (illustrated in the large panel of Figure 1.2) closely mirrors low compliance-cost type plants. The last panel of Figure 1.2 shows high compliance-cost plants undertaking more abatement after the regime change, just as the model anticipated. These effects only temper aggregate behavior however, because of the comparatively small share of the population of Chilean plants that are high compliance-cost types.

1.6.1 Main Results

The central conclusions of the graphs in Figure 1.2 warrant a more quantitatively rigorous investigation via regression analysis. The results of regression estimates following the specifications of Equations 1.9-1.12 are presented in Table 1.2. We see that the simplistic comparison of abatement in 2000 to 2001 yields an estimated 5.8 percentage point reduction in abatement across all plants in response to the change in performance standards. This is from an average abatement of 19.2% under the Relative regime, suggesting an aggregate reduction in abatement of 30.3% as a result of the switch from Relative to Absolute. This aggregate reduction is the result of estimated effects of a 13.5 percentage point reduction and a 27.9 percentage point increase in abatement for plants of low and high compliance-cost types respectively.

To focus the analysis more closely on the effects of the regime change, the regressions are also estimated on the longer (1995-2010) sample with a full set of plant-specific controls and pre and post time trends. This second approach (as detailed in Equations 1.11-1.12) is my preferred analysis, the results of which are presented in the last two columns of Table 1.2 and closely mirror estimates from the first analysis. In aggregate, we see a nearly 6.4 percentage point reduction in estimated abatement in response to the change in performance standard. Just as we expected from our conceptual discussion, the sign of this estimate is driven by the behavior of low compliance-cost plants, which are estimated to have reduced their abatement by 14.1 percentage points in response to the standard change. The magnitude of aggregate abatement is tempered however by the response of the high compliance-cost plants, which increase their abatement levels by an average of 27.7 percentage points as a result of the change in performance standards.[28]

The substance of these results is also reflected in Table 1.3, which presents the results of the main specifications with Abater and De-Abater status as the outcome variables. These estimates suggest that overall, plants are less likely to undertake abatement under the Absolute than Relative performance standard, and that de-abatement (i.e., a year-on-year *increase* in concentration) is more likely after the regime change. Again, these aggregate results are attributable to heterogeneous effects between plants with low

[28]As 2001 was the first year under the new performance standard it may be that it served as an adjustment year when the full implications of the new regime had not yet been internalized. However, dropping 2001 from the analysis (and thereby comparing abatement in 2000 to abatement from 2000 to 2002) yields estimates that are very similar to those in my main analysis (results not shown).

and high compliance-costs. Just as low compliance-cost plants undertake less abatement on average after the regime change, they are also less likely to undertake any abatement and more likely to de-abate. The reverse of each of these results is true for plants of high compliance-cost type, identifying again the sharply heterogeneous effects anticipated by the conceptual model.

Taken together, the results presented thus far tell a compelling story of disparate responses to the change in performance standards. Plants with low compliance-costs under the examined threshold, undertake lower levels of abatement after the regime change than before, suggesting that the Relative regime was more effective in motivating abatement among this group. Conversely, high compliance-cost plants undertook more abatement after the change in performance standard, implying that the Absolute standard better incentivized this group to abate. These results embody heterogeneous comparative efficacies of the Relative and Absolute performance standards depending on individual costs-of-compliance with the punishment threshold. Additionally, these dichotomous effects strongly demonstrate the dependence of aggregate compliance efforts on the distribution of these agent types within the regulated population. Thus, we see the strong population majority of low compliance-cost types resulting in the Relative regime driving more abatement on average than did the Absolute performance standard.

1.6.2 Policy Assessment - Weighted Results

While the previous results examined the impacts of the regime change on compliance decisions, the contribution of the emissions cleanup program to overall air quality in Santiago is also of interest. Because a given level of abatement from a small operation will not reduce the total amount of particulate matter emitted by as much as the same level of abatement undertaken by a large emitter (due to the difference in emissions volumes between the two plants), I undertake a weighted regression analysis in addition to my main analysis.

To capture the effect of the regime change on total emissions of particulate matter into the atmosphere, my preferred regression specifications are re-estimated for abatement, Abaters, and De-Abaters with weights assigned according to the daily emissions volume of each plant. The results of this analysis are presented in Table 1.4, and suggest that the change in performance standard from Relative to Absolute in 2001 led to a large slowdown in the cleanup of emissions. These results continue to be driven by the plants with low compliance-costs, and the fact that the estimates are even more disparate than those presented previously suggests that large plants are reacting to the change in performance standard more strongly than small plants. This characterization does not carry over to the weighted De-Abater analysis which is insignificant in aggregated and for low compliance-cost plants, suggesting that while large emitters with low compliance-costs undertook less abatement under the Absolute regime than the Relative, they were unlikely to actually increase their concentrations (i.e., de-abate) in response to the standard change. As such, the significant coefficients on De-Abater in the unweighted analysis must be driven by a higher likelihood that smaller plants undertake de-abatement in response to the standard change.

In total, the weighted regression results suggest that the 2001 change from the Relative to the Absolute performance standard made the emissions cleanup program less effective in reducing total emissions of particulate matter. Additionally, these results suggest that in the context of the Chilean emissions cleanup program, the Relative regime was a more effective means of inducing aggregate reductions in the total emissions of particulates than was the Absolute regime.

1.7 Robustness Checks

Having outlined my main results, I now address a number of potential issues that could impact the interpretation or validity of these analyses. I first show that plants change their behavior in response to punishments and then empirically validate the approach used to identify plants as having either low or high compliance-costs. Next, I demonstrate that the adoption of cleaner technologies was neither the means nor driver of the changes in abatement we identify between 2000 and 2001. A placebo test is presented showing that my method of identification does not find effects on emissions characteristics for which changes were not incentivized under the emissions cleanup program. Finally, I show that my results are not attributable to systematic plant exits.

The same punishments are the basis of both the Relative and Absolute performance standards in Chile. If these punishments are ineffectual, the means by which they are assigned is inconsequential. I therefore empirically confirm that the punishments in the Chilean context did in fact impact abatement decisions. Table 1.5 presents the results from a regression of abatement levels and Abater status on an indicator for punishment assignment in the previous year and a count of Episodes in the previous year. The results suggest that higher levels of abatement and a higher probability of undertaking any level of abatement are associated with punishment in the prior year, and higher probabilities and levels of abatement were observed across all plants when the most recent punishments were more costly. These estimates suggest generally that plants are responsive to the assignment and costs of punishment.[29]

The empirical results regarding heterogeneous plant behavior depend on the separation of plants into the low and high compliance-cost types which arose from the conceptual model. Figure 1.3 shows that there is a meaningful change in comparative behavioral responses to the two performance stan-

[29]Note that abatement levels do not predict the number of Episodes in the coming year, suggesting that plants cannot effectively identify (or cause) years in which few Episodes are likely to be announced and adjust their behaviors accordingly (results not shown).

dards at (or just slightly below) the cut-point used to map plants empirically to low and high compliance-cost groups. This non-parametric characterization of abatement under each regime demonstrates that the bin differentiation suggested by the conceptual model corresponds very closely to a meaningful difference in behavioral responses seen in the data.

Additionally, we can look at Figure 1.3 as a noisy empirical reflection of the relationships outlined conceptually in Figure 1.1. Although the mapping between compliance-costs and year 2000 operational concentrations is only exact for dividing plants between types, both compliance-costs and year 2000 operational concentrations are increasing in baseline concentrations. Thus we might treat the observed year 2000 concentrations as a loose proxy for compliance-costs and see how the realized changes in abatement (the distance between the lines in Figure 1.3) matches the predictions as graphed in Panel B of Figure 1.1. As already noted, the cross point of empirical abatement levels under the two regimes occurs at almost exactly the predicted level. We also see that the signs of the difference in abatement match those predicted by the model both above and below the crossing point, and that for plants with higher year 2000 concentrations, the levels of abatement under the two regimes converge and become statistically indistinguishable (although the density of plants is quite low at high concentrations). Finally, although we do not empirically observe the convergence in behavior predicted by the model at low concentration levels, this is likely due to the poor mapping between baseline concentrations and operational concentrations at low levels. Taken as a whole, the non-parametric representation of plant behavior in Figure 1.3 demonstrates the empirical validity of the predictions of the conceptual model and the proposed type-mapping procedure.

Because I have assumed in the conceptual model that technology was fixed over the time of the regime change, the interpretation of my results and the conclusions from the model would be undermined if adoption of clean technology was the means by which abatement shifted as a result of the standard change. Additionally, the identified results could be confounded if adoption of cleaner technology occurred differentially between 2000 and 2001, but not in response to the change in performance standards. Natural gas was the main option for those plants wanting to adopt a cleaner and/or cheaper technology during the study period. Table 1.6 reports the results of a hazard model of natural gas adoption assessed over both the 2000-2001 sample and the 1995-2010 sample. A plant is considered "at-risk" (and therefore enters the sample) when natural gas becomes available in its neighborhood of operation and is no longer "at-risk" if natural gas is adopted.[30] The first two columns of results in Table 1.6 demonstrate that there was a change in natural gas adoption rates contemporaneous to the change in performance standards, but we see in the last two columns of the table that once we control for other factors, most notably the cost ratio of natural gas to the current fuel of each plant, there is no longer a significant relationship between the causal factor of interest in this investigation (the regime change) and the potentially confounding means of abatement (natural gas adoption).[31] As such, the exclusion of technology change from consideration is not likely to have impacted my results or interpretation, and suggests that major investments are not undertaken to address uncertainty surrounding threshold levels, but instead to address larger, more enduring changes to regulatory or cost structures.[32] These findings match the conclusions of Coria (2009), which found that the price differential between natural gas and a plant's current fuel drove the decisions of Chilean plants to switch to natural gas, with environmental regulations having very little impact

[30]Natural gas first became available in Santiago in 1997, and service spread across the city in subsequent years. I define the year of initial availability for a neighborhood as the first year in which any plant in the neighborhood used natural gas.

[31]The natural gas to current-fuel-price ratios are not included in the mainline specification, but their inclusion does not markedly impact the results. This control was left out of the main analysis because the cost ratios are not available for the whole time period of interest or for all fuels used by plants in the data.

[32]For instance, the persistent lower costs of natural gas, or the regularly falling threshold level under the Relative regime. The ratcheting effect of a multi-period Relative standard is not examined in this paper, as I have sought to address the comparative effectiveness of the Absolute and Relative regimes under a fixed punishment threshold.

(similar results are also reported by: Montero et al., 2002).

To ensure that the identified effects in abatement do not also exist in other emissions measures which were not addressed by the emissions cleanup program, I re-estimate the main regressions using changes in emissions flow as an outcome variable.[33] Total particulate matter emissions can be cut back either by reducing the concentration of particulate matter in emissions or by reducing the volume of emissions (i.e., flow), however, only changes in concentration were incentivized via the emissions cleanup program examined in this paper. Table 1.7 demonstrates that, with full controls, flow was unaffected by the regime change, suggesting that we have been examining the right measure of impact and have not been picking up effects driven by some alternative program or factor outside the change in performance standards.

Having determined that the change in performance standard appears to be impacting reasonable emissions characteristics, I now consider whether the main results of this paper could be due to sample selection as a result of selective plant exits from the data. Table 1.8 presents the 95% confidence intervals for the short sample (i.e., 2000-2001) effect of the regime change by low and high compliance-cost types, following the bounding procedure laid out by Lee (2009). This so-called "Lee Bounds" procedure trims the year 2000 sample to match the sample that remains in 2001, by cutting the most extremely impacted observations from one side of the distribution. These results can be characterized as "the tightest bounds for the average treatment effect that are consistent with the observed data" (Lee, 2009), and therefore it is comforting that the main thrust of my results remains intact for this somewhat forgiving bounding approach.

The results are also robust to a more conservative "bounding" proce-

[33]The diluting of emissions is closely regulated and monitored. As a result, the levels of flow and concentration are kept independent by government regulation and oversight. Ensuring that this separation is reflected in the data, I find no predictive power of changes in concentration on either the levels or changes of flow (results not show), and thus I treat flow as an independent emissions parameter which is unaddressed by the emissions cleanup program.

dure in which all plants that were active in 2000, but not in 2001, are assigned 2001 counterfactual abatement levels of the 5th and then 95th percentile levels of abatement by plant type. The results of this bounding exercise are presented in Table 1.9, which shows again that my central results are not being driven by selection in exit behavior by plants.

1.8 Changing the Distribution of Types

My conceptual model highlights the distribution of plant types as central to the comparison of performance standards, and empirically I find that the aggregate results are driven by the population balance of low and high compliance-cost types. I now present an illustrative example of how the comparison of performance standards is impacted by the distribution of plant types.

Empirically, I was only able to examine the single distribution of Chilean plants as it was observed at the time of the Chilean change in performance standards. In order to assess how my results might vary under different distributions of plant types, I simulate a new population of plants by stratifying the empirical sample according to compliance-cost types, and randomly drawing a sub-sample of the population with an increased share of high compliance-cost plants, which is akin to simulating a stricter incentive framework.[34] Specifically, 1,253 (of 3,132) low compliance-cost plants and 626 high compliance-cost plants (i.e., all of them) are drawn from the full sample without replacement to create the simulated sample. This yields a simulated population in which the share of high compliance-cost plants is double that of the original sample (33.2% versus 16.6%) at the time when the change in performance standards occurred.

Table 1.10 presents the results of my main specifications estimated on

[34]Note that the strictness of a given threshold can be altered either by moving the threshold for a given population or by "moving" the population relative to the threshold. I have followed the latter approach here.

both the empirical sample (reproducing results already presented in Table 1.2) and the simulated sample. Comparing these results reveals essentially identical estimates for the heterogeneous effects among the low and high compliance-cost groups, which suggests that the behavior of the two types in the simulated sample is representative of that in the real-world population. The aggregate effect of the policy change is however dramatically different for the simulated sample compared to that for the empirical population. The Absolute regime was shown to be less effective than the Relative regime in motivating abatement in the main analysis, but the estimated effect of the regime change on the simulated sample is a statistical zero. This analysis was iterated 100 times, with a different simulated population each time. Similar to the results reported in Table 1.10, the average magnitude of the aggregate effect for the fully controlled, long-sample analysis is about 0.1 percentage points (with a comparatively large average standard error of 1.636 percentage points) while the separate average effects for the low and high compliance-cost groups are -15.3 percentage points and 27.6 percentage points respectively, and both are always significant at the 0.01 level. Just as we would have anticipated from the predictions of the conceptual model, this exercise demonstrates that an increase in the strictness of regulation leads to improved effectiveness of the Absolute standard in comparison to the Relative performance standard due to a shift in the balance of low and high compliance-cost types in the population.

1.9 Conclusion

Absolute and Relative performance standards serve as alternative approaches to incentive allocation in a wide range of settings, and although these standards have been the subject of much theoretic inquiry, causal empirical comparisons have proven elusive. In this paper, I take advantage of a natural experiment that changed only the regulatory performance standards between two implementations. The repeated nature of the incentivized task allows me to properly control for task and agent specific factors, and the observability of compliance behaviors under the emissions cleanup program allows us to draw straightforward comparisons between the effectiveness of Absolute and Relative performance standards in motivating effort.

Through the analysis of this paper I seek to address two distinct, but related questions. First, “Does an Absolute performance standard or a Relative performance standard motivate more effort, and what conditions impact this comparison?” and second, “In Chile, what was the impact of the 2001 change from a Relative to Absolute standard on year-on-year improvements in air quality?”. The first question treats each firm as a decision maker of equal interest to regulators (as one might expect a teacher to value effort from all students equally), while the latter question serves as a policy assessment for the change in performance standards by placing more weight on the behavioral changes of firms controlling larger plants whose emissions contribute more to overall air quality.

In the more general examination of plants as decisions makers, I find that the Relative standard was 6.4 percentage points more effective than the Absolute standard in motivating effort among Chilean plants. Perhaps more importantly, I find that such aggregate effects are attributable to the population balance between agents with heterogeneous comparative responses to the two standards. Plants with low compliance-costs are found to undertake 14 percentage points less abatement under the Absolute regime, while plants with high compliance-costs are found to undertake nearly 28 percentage points *more* abatement under the Absolute standard as compared to the Relative.

My assessment of the comparative contribution of the emissions cleanup program under the Relative and Absolute performance standards to the improvement of Santiago’s air quality is of particular interest, given the difficulty of regulating atmospheric pollution in developing economies (see for example: Duflo et al., 2013; Ghanem and Zhang, 2014; Greenstone et al., 2011; and Pande et al., 2012). Weighting plants according to their size reveals that the

2001 standard change resulted in a 14.5 percentage point reduction in the level of annual improvement to the total emissions of particulate matter. This suggests that the emissions cleanup program was significantly better supported by the Relative regime.

Finally, I demonstrate the sensitivity of the aggregate results to the distribution of plant types by increasing the share of high compliance-cost plants in a simulated sample population. This is akin to an increase in regulatory strictness since the average cost of regulatory compliance within the population is raised. I show that the stricter a regulation becomes, the better an Absolute regime preforms compared to a Relative, demonstrating once again how important the balance of types is in the comparison of performance standards.

While the particulars of other settings will never exactly match those of the Chilean emissions cleanup program, the general conclusions of this paper provide valuable insight for any principal seeking to motivate heterogeneous agents. Broadly I show that the same policy can be much more effective when enforced by one standard versus another, and that the more effective standard in a given setting is determined by the distribution of low and high compliance-cost types in the relevant population. When the balance is tipped toward agents with high compliance-costs (i.e., the regulation becomes stricter) the comparative effectiveness of an Absolute regime improves, while laxer regulation improves the performance of a Relative regime compared to an Absolute one. Taken together, my results underscore the importance of fitting the performance standard of any incentive framework to the population and the distribution of agent types therein.

It is perhaps instructive to think about my results in the familiar setting of classroom grading standards. My findings suggest that a curved grading scheme may not motivate low ability students as well as an Absolute scheme because the additional uncertainty between effort and grades serves as an extra hurdle that must be overcome. The uncertainty will cause some low ability students (i.e.- high compliance-cost types) to "give up" and others to miss higher grades they might have been willing to work for had the threshold been known. Conversely, the uncertainty of a curve motivates students with high abilities (i.e.- low compliance-cost types) to perform above the relevant threshold as a means of ensuring the high grade when the exact level of the threshold is not known ex ante. Intuitively, the number of low-ability students that "give up" under the curved grading scheme grows as the grading standard is raised (i.e., made stricter) and a higher share of students fall into the low-ability/high compliance-cost category for the course.

This paper has examined the effects of uncertainty in the level of the regulatory threshold on plant compliance efforts assuming that the exact operational concentration can be chosen and achieved each period. In practice, it is likely that plants cannot precisely select the concentration measured by auditors for the upcoming regulatory cycle, but instead experience some variability in their realized levels of concentration.[35] This additional degree of uncertainty undoubtedly interacts with the threshold uncertainty in plant decisions, providing an interesting avenue for further investigation.

1.10 Acknowledgements

Chapter 1, in full, is currently under review for publication. Jamie Mullins, the dissertation author, was the primary investigator and author of this paper.

[35]This is the type of variability investigated by Bandyopadhyay and Horowitz (2006) and Shimshack and Ward (2008).

1.11 Tables and Figures

Table 1.1: Summary Statistics by Year

Year	**Plants**	**Total TME** (kg/hr)	**Mean TME** (kg/hr)	**Mean Conc.** (mg/m^3)	**Mean Flow** (m^3/hr)	**Mean Abatement** (mg/m^3)	**Abtrs**	**De-Abtrs**	**Eps.**
1995	3,484	668.4	0.1925	67.4	3,033.2	-	-	-	2
1996	3,671	706.0	0.1923	62.6	3,034.1	3.3	523	393	6
1997	4,015	395.9	0.0986	46.2	2,515.2	5.3	428	372	13
1998	4,105	369.2	0.0905	44.8	2,596.6	1.4	736	804	14
1999	3,989	260.7	0.0654	31.4	2,947.5	12.4	2,076	266	16
2000	4,362	208.4	0.0478	25.1	2,969.0	4.7	1,614	473	11
2001	4,890	182.9	0.0374	21.7	2,715.4	2.8	1,475	515	4
2002	5,121	164.6	0.0322	19.9	2,664.2	1.3	1,260	815	11
2003	5,432	176.6	0.0325	19.8	2,580.1	0.0	1,031	1,116	5
2004	5,641	168.1	0.0298	18.0	2,488.3	1.7	2,664	593	2
2005	6,423	181.4	0.0282	17.7	2,353.7	0.1	553	453	2
2006	6,948	204.3	0.0294	17.6	2,262.1	0.0	550	474	3
2007	7,103	188.8	0.0266	17.5	2,168.1	0.0	520	593	4
2008	7,850	236.6	0.0302	17.5	2,176.7	-0.2	513	582	8
2009	8,527	227.8	0.0267	17.3	2,100.2	0.0	556	516	0
2010	8,804	197.5	0.0224	16.9	2,013.0	0.2	597	519	2

Notes: Summary statistics include all observations for which information on emissions flow and concentration is provided for two consecutive years except “generator” type plants, which do not enter the data until 2004 and are thus excluded from this examination. The data from 1995 is considered correct but incomplete, as not all active plants were cataloged. Additionally, the Environmental Episode program was not consistently enforced until 1997, and thus, the small number of Episodes in 1995 and 1996 is a reflection of the under-use of the Episode policy rather than an indication of good air quality (See Mullins and Bharadwaj, 2014 for further discussion). Episode counts include only Pre-Emergency and Emergency level Episodes. TME - Total Mass Emissions - is the product of flow and concentration, both of which are measured at full operational capacity following the procedures of Environmental Protection Agency, Method 5. Abtr is short for Abater which is any plant that reports a lower concentration in year t than in year $t-1$, conversely, De-Abtr is short for De-Abater which is any plant that reports an increased concentration level in year t compared to year $t-1$.

Panel A. OVFs and Bin Mappings with assumed *Beta(3,2.5)* Threshold Distribution

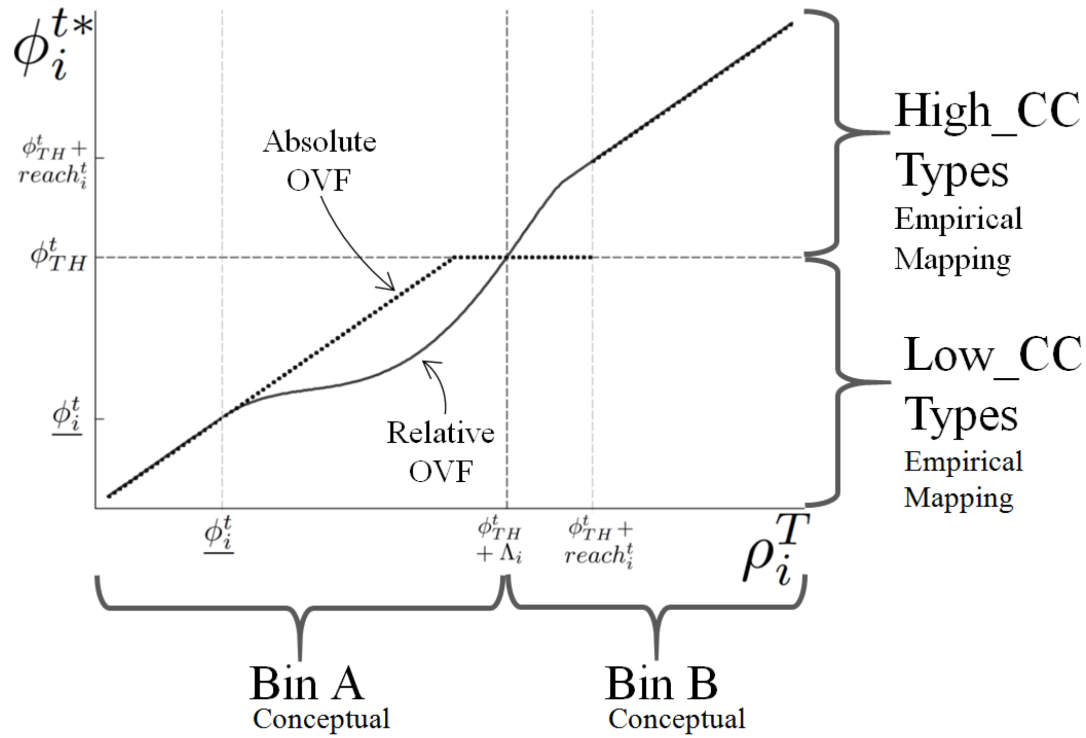


Figure 1.1: Type Mappings & OVF under Absolute and Relative Regimes: $Beta(3, 2.5)$ Threshold Distributions

Panel B. Predicted Change in Abatement
in Response to the Change in Performance Standards

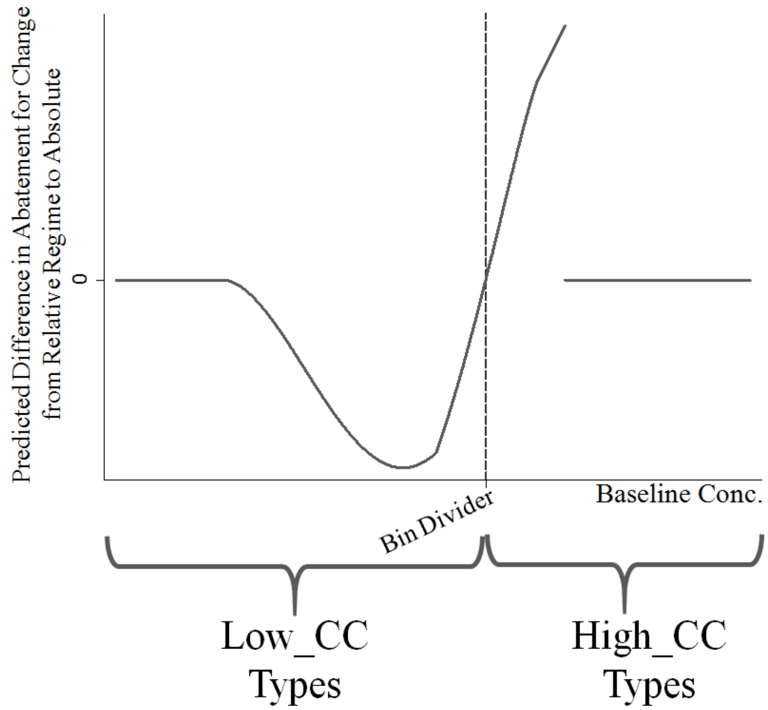


Figure 1.1: Type Mappings & OVF under Absolute and Relative Regimes: $Beta(3, 2.5)$ Threshold Distributions - Continued

Notes: ρ_i is baseline concentration and ϕ_i^{t*} is the optimal concentration of plant i in period t. The dotted line in Panel A represents optimal concentrations under the Absolute regime. Low_CC and High_CC are low and high compliance-cost types respectively. The solid line in Panel A represents optimal concentrations under the Relative regime for a plant that believes the threshold level is distributed $beta(3, 2.5)$ on the interval $\left[\underline{\phi_i^t}, \overline{\phi_i^t}\right]$. This distributional assumption is made for illustrative purposes. The type assignment according to the empirical mapping is determined by the level of the Relative OVF compared with the threshold level, ϕ_{TH}^t, because the point at which $OVF^R = OVF^A$ is always the same as the point at which $OVF^R = \phi_{TH}^t$. Graphs are intended to depict conceptual relationships and are not based on empirical data.

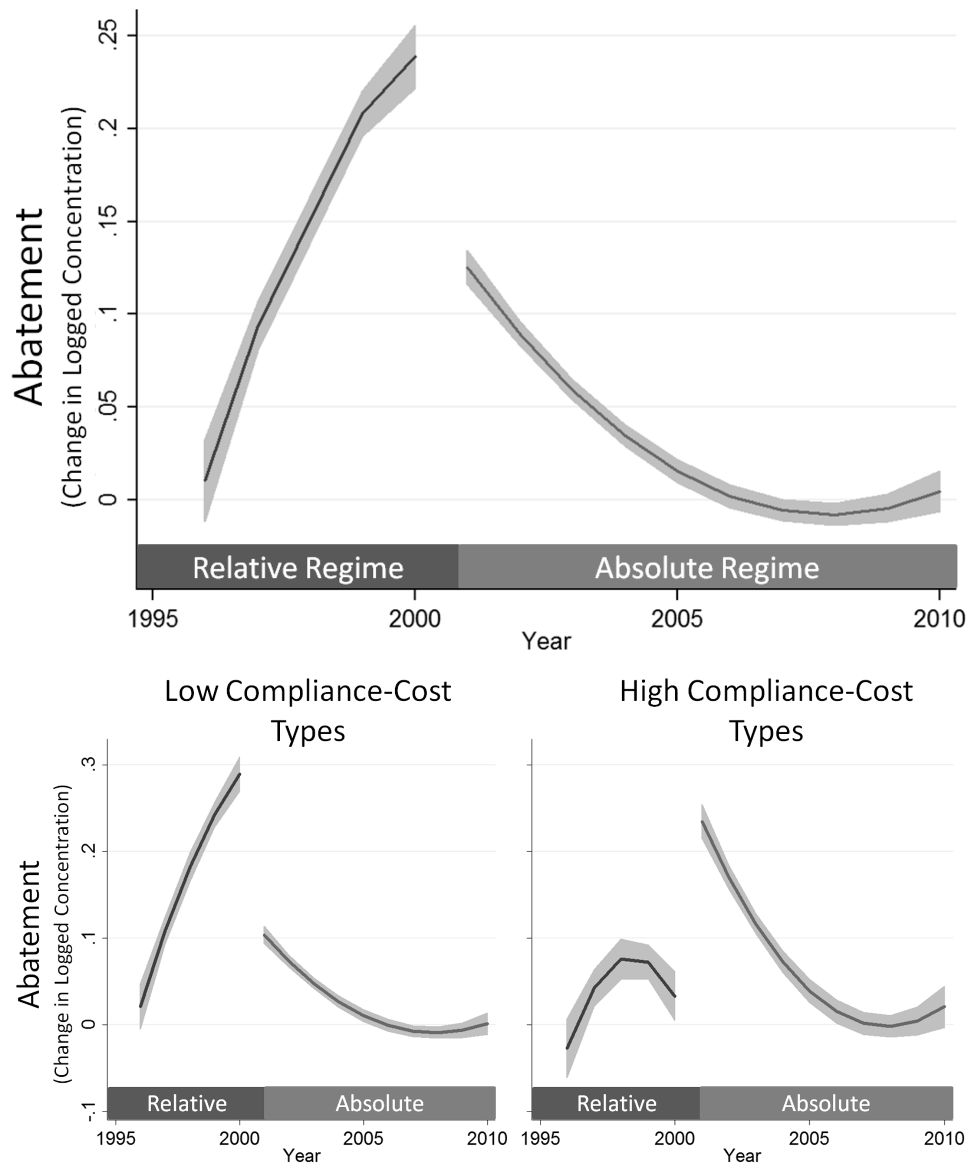


Figure 1.2: Mean Abatement by Year with 95% Confidence Intervals

Notes: Plotted curves are fractional polynomial fits of abatement (change in logged concentration) by year. Graphs include all observations for the 3,758 plants for which abatement could be calculated in 2001. Gray shading represents 95% confidence intervals for the smoothed annual mean abatement levels.

Table 1.2: Main Results - Abatement by Regime and Compliance-Cost Types

Plant Counts in 2001	VARIABLES	2000-2001 Change in Logged Conc.	2000-2001 Change in Logged Conc.	1995-2010 Change in Logged Conc.	1995-2010 Change in Logged Conc.
3,758	APS (0/1)	-0.0582***		-0.0635***	
		(0.01)		(0.01)	
3,132	APS * Low_CC		-0.135***		-0.141***
			(0.01)		(0.01)
626	APS * High_CC		0.279***		0.277***
			(0.03)		(0.03)
Year 2000 Abatement		0.192			
Quadratic Time Trends		No	No	Yes	Yes
Additional Controls		No	No	Yes	Yes
SE Clustered by Firm		Yes	Yes	Yes	Yes
Observations		6,663	6,663	39,688	39,688
Number of Clusters		2,384	2,384	2,384	2,384
R-squared		0.005	0.183	0.088	0.127

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: Analysis includes all observations for which non-zero flow values are reported and abatement can be calculated in 2001. APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

Table 1.3: Abater/De-Abater Results: 1995-2010 Sample

VARIABLES	Prob. Abater	Prob. Abater	Prob. De-Abater	Prob. De-Abater
APS (0/1)	-0.278***		0.290***	
	(0.04)		(0.05)	
APS *Low_CC		-0.486***		0.498***
		(0.05)		(0.06)
APS * High_CC		0.471***		-0.315***
		(0.11)		(0.11)
Quadratic Time Trends	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes
SE Clustered by Firm	Yes	Yes	Yes	Yes
Probit	Yes	Yes	Yes	Yes
Observations	39,688	39,688	39,688	39,688
Number of Clusters	2,384	2,384	2,384	2,384

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: Probit model is used. An Abater is any plant that reports a lower concentration in year t than in year $t-1$, conversely, a De-Abater is any plant that reports an increased concentration level in year t compared to year $t-1$. Analysis includes all observations for which non-zero flow values are reported and abatement can be calculated in 2001. APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

Table 1.4: Results Weighted by Daily Flow Rate: 1995-2010 Sample

VARIABLES	Change in Logged Conc.	Change in Logged Conc.	Prob. Abater	Prob. Abater	Prob. De-Abater	Prob. De-Abater
APS (0/1)	-0.145**		-0.286***		0.0246	
	(0.06)		(0.07)		(0.06)	
APS * Low_CC		-0.207***		-0.324***		0.053
		(0.06)		(0.07)		(0.07)
APS * High_CC		1.060***		0.433***		-0.498***
		(0.23)		(0.08)		(0.09)
Year 2000 Weighted Mean	0.352		0.672		0.159	
Quadratic Time Trends	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
SE Clustered by Firm	Yes	Yes	Yes	Yes	Yes	Yes
Linear Probability Model	No	No	Yes	Yes	Yes	Yes
Observations	39,625	39,625	39,625	39,625	39,625	39,625
Number of Clusters	2,384	2,384	2,384	2,384	2,384	2,384
R-Squared	0.094	0.118	0.11	0.492	0.063	0.337

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: Analytical observation weights equal to the product of the reported flow and daily operational hours are used. Analysis includes all observations for which abatement can be calculated for 2001, and for which non-zero flow and operational hours are reported. There are 63 fewer observations in this analysis than in the mainline analysis due to 63 observations reporting zero hours of daily operation. An Abater is any plant that reports a lower concentration in year t than in year $t-1$, conversely, a De-Abater is any plant that reports an increased concentration level in year t compared to year $t-1$. Linear Probability Models were used for the Abater and De-Abater regressions rather than the Probit to better accommodate the observation weights. APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

Table 1.5: Incentive Relevance

	Abatement: 1995-2010		Abater: 1995-2010	
VARIABLES	Change in Logged Conc.	Change in Logged Conc.	Prob. Abater	Prob. Abater
# of Episodes in Prior Year	0.00956***	0.00896***	0.0223***	0.0154***
	(0.000469)	(0.000484)	(0.000546)	(0.000551)
Punished in Prior Year	0.247***	0.198***	0.209***	0.155***
	(0.0118)	(0.0115)	(0.00974)	(0.0102)
Additional Controls	No	Yes	No	Yes
SE Clustered by Firm	Yes	Yes	Yes	Yes
Linear Probability Model	No	No	Yes	Yes
Observations	39,688	39,688	39,688	39,688
Number of Clusters	2,384	2,384	2,384	2,384
R-squared	0.055	0.094	0.096	0.2

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: "# of Episodes in Prior Year" is simply a count (constant across all plants) of Episodes announced in the previous year, while "Punished in Prior Year" is a plant-specific indicator variable for whether the plant was assigned to the Shutdown List in the previous year. Analysis includes all observations for which non-zero flow values are reported and abatement can be calculated in 2001. An Abater is any plant that reports a lower concentration in year t than in year $t-1$. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

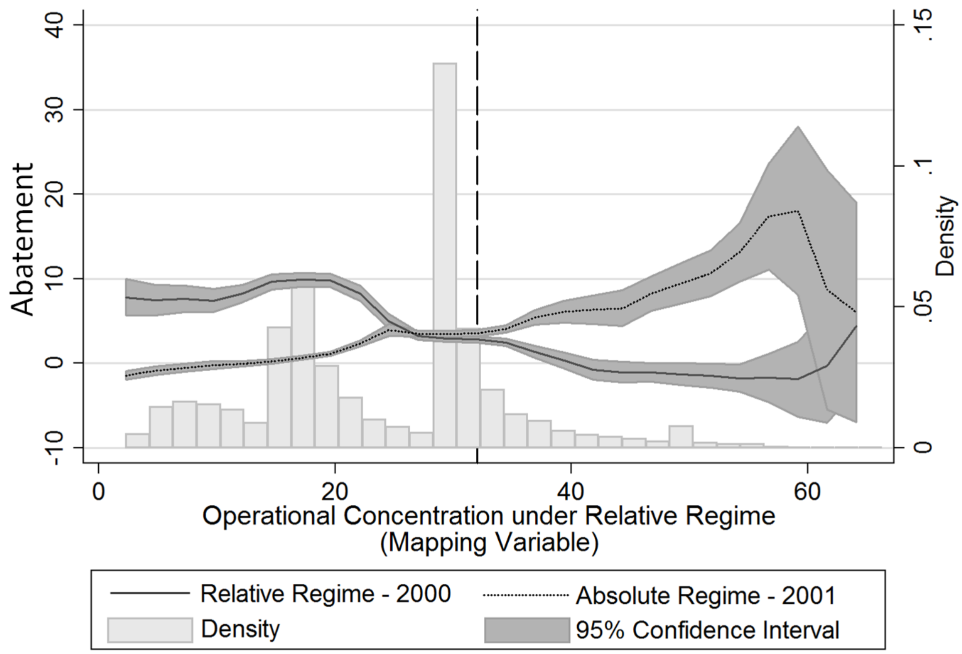


Figure 1.3: Abatement by Pre-Regime-Change Operational Concentration

Notes: Lines represent polynomial smooths of abatement by year 2000 operational concentration. Graph includes all plants in main analysis. The vertical dashed line at $32\mu g/m^3$ indicates the operational concentration under the Relative regime used maps plants into Bins A and B empirically (i.e., low and high compliance-cost types). This division point is based on the empirical mapping of differential behavioral predictions of the regime change from the conceptual model.

Source: Elaborated from PROCEFF Data.

Table 1.6: Adoption of Natural Gas: Probit Hazard Model

VARIABLES	2000-2001 Probability Adopt NG	 Prob. Adopt NG	1998-2005 Prob. Adopt NG	 Probability Adopt NG
APS (0/1)	-0.0710***		0.00149	
	(0.01)		(0.06)	
APS * Low_CC		-0.124***		0.00243
		(0.02)		(0.06)
APS * High_CC		0.153***		-0.0713
		(0.02)		(0.07)
Cost Ratio of NG to Current Fuel			-0.522***	-0.524***
			(0.02)	(0.02)
Quadratic Event Time Trend	No	No	Yes	Yes
Additional Controls	No	No	Yes	Yes
Probit	Yes	Yes	Yes	Yes
SE Clustered by Firm	Yes	Yes	Yes	Yes
Observations	5,941	5,941	13,873	13,873
Number of Clusters	2,189	2,189	2,130	2,130
R-squared	0.009	0.194	0.490	0.549

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: Cost ratios are taken from (Coria, 2007), and are only available for 1998-2005, limiting the time-frame of these analyses. Event time rather than year were used in these Hazard regressions, with $t = 0$ set at the year in which natural gas became available in a plant's neighborhood. Availability is defined as the year in which any plant in a given neighborhood first adopted natural gas. Note that observations only enter the analyzed sample once natural gas becomes available in their neighborhood, and they exit the sample if and when they adopt natural gas. Some plants never had natural gas as an option and some used fuels for which cost ratios are unavailable. In either case, such plants are not included in this analysis. APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls are as described previously.

Table 1.7: Placebo Test - Flow as Outcome

	2000-2001		1995-2010	
VARIABLES	Change in Logged Flow	Change in Logged Flow	Change in Logged Flow	Change in Logged Flow
APS (0/1)	-0.0281**		0.0107	
	(0.01)		(0.01)	
APS * Low_CC		-0.0284*		0.00945
		(0.02)		(0.02)
APS * High_CC		-0.0313		0.0125
		(0.02)		(0.03)
Quadratic Time Trends	No	No	Yes	Yes
Additional Controls	No	No	Yes	Yes
SE Clustered by Firm	Yes	Yes	Yes	Yes
Observations	6,663	6,663	39,688	39,688
Number of Clusters	2,384	2,384	2,384	2,384
Pseudo R-squared	0.001	0.004	0.009	0.012

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: Flow is a measure of maximum plant emissions volume. The level of flow was unregulated during the study period, however “excess air” regulations (which were unchanged throughout the study period) ensured that concentration and flow levels were set independently. Thus changes in flow were not incentivized (even indirectly) under either the Absolute or Relative regimes. As such we don’t expect the switch in performance standards to materially impact flow. Analysis includes all observations for which non-zero flow values are reported and abatement can be calculated in 2001. APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

Table 1.8: Lee Bounds on the Impact of Exiter Sample Selection: 2000-2001 Sample

95% Confidence Interval for Abatement Effect of Regime Change by Compliance-Cost Type		
	Lower	Upper
Low_CC	-0.2611	-0.0324
High_CC	0.1931	0.3289

Notes: Bounds are "tightened" when intra-type variation exists using indicator variables for natural gas as an abatement option. Bounding procedure follows Lee (2009). Analysis includes all observations for which non-zero flow values are reported and abatement can be calculated in 2001. Low_CC and High_CC are indicators for low and high compliance-cost types respectively.

Table 1.9: Bounding Results for 2001 Exiters: 1995-2010 Sample

	2001 Exiters Assigned 5th Percentile Abatement Level		2001 Exiters Assigned 95th Percentile Abatement Level	
VARIABLES	Change in Logged Conc.	Change in Logged Conc.	Change in Logged Conc.	Change in Logged Conc.
APS (0/1)	-0.0857***		0.00231	
	(0.01)		(0.01)	
APS * Low_CC		-0.159***		-0.0787***
		(0.01)		(0.01)
APS * High_CC		0.234***		0.358***
		(0.03)		(0.03)
Quadratic Time Trends	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes
SE Clustered by Firm	Yes	Yes	Yes	Yes
Observations	41,105	41,105	41,105	41,105
Number of Clusters	2,631	2,631	2,631	2,631
R-Squared	0.086	0.124	0.091	0.135

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: 2001 exiters are those that appear in the 2000 data but are absent thereafter. 365 plants are 2001 exiters. These are added back in for the bounding exercise and assigned first the 5th and then the 95th percentile abatement level in 2001 for the compliance-cost type in which each was a member. In addition to the 2001 observations added for these plants for the analyses in columns 3-6, all prior observations of these plants are also included (hence the observation count grows by more than 365). APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

Table 1.10: Simulated Sample versus Empirical Sample

# in Simulated Sample	VARIABLES	Empirical Sample 1995-2010 Change in Logged Conc.	Empirical Sample 1995-2010 Change in Logged Conc.	Simulated Sample 1995-2010 Change in Logged Conc.	Simulated Sample 1995-2010 Change in Logged Conc.
1,879	APS (0/1)	-0.0635***		0.00141	
		(0.01)		(0.02)	
1,253	APS * Low_CC		-0.141***		-0.153***
			(0.01)		(0.02)
626	APS * High_CC		0.277***		0.276***
			(0.03)		(0.03)
Quadratic Time Trends		Yes	Yes	Yes	Yes
Additional Controls		Yes	Yes	Yes	Yes
SE Clustered by Firm		Yes	Yes	Yes	Yes
Observations		39,688	39,688	20,063	20,063
Number of Clusters		2,384	2,384	1,439	1,439
R-squared		0.088	0.127	0.082	0.13

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Robust Standard Errors Clustered by Firm in Parenthesis ( )

Notes: Compared to the empirical population, the simulated population has double the share of high compliance-cost plants (33.2% versus 16.6%). The simulated population is generated by randomly dropping 1,879 low compliance-cost plants. APS is an indicator for the use of the Absolute performance standard. Low_CC and High_CC are indicators for low and high compliance-cost types respectively. Additional Controls include initial levels of flow and fuel consumption, an indicator for whether switching to natural gas was an option, an indicator for whether a plant participated in a Clean Fuel Program, and indicators for whether a plant began the period in violation of concentration ceilings. Each firm may control one or multiple plants.

1.12 Appendix Tables

Table 1.11: Yearly Timeline of Measurement Submission and List Publication

Date	Event
Prior to February	Plants have the opportunity to set concentration for year, submit to third party measurement of emissions, and provide measurements to Chilean government.
February	Santiago Metropolitan Region's Health Services Agency (SESMA) compiles preliminary ranked and Shut Down lists. The preliminary Shutdown List is published, and plants on the Shut Down list are notified via certified mail of their status. Plants have 10 days from receipt of the certified notification to challenge assignment to the Shutdown List
March	The final Shutdown List is published online and in nationally and regionally circulating newspapers.
April-August	Episode program is in effect, and Episodes may be announced in order to address realized or expected poor air quality. Upon Episode announcement of either Pre-Emergency or Emergency levels, plant shutdowns are required.

Notes: There is some variation in timing over the period of the study with list publication happening earlier in the later years of my sample. The above timeline is presented as an example and represents the schedule in 2001, as outlined in an update to the original law that set up the Environmental Episode program, available: http://www.leychile.cl/Navegar?idNorma=7871

Table 1.12: Episode Levels and Protocols in 1999

Episode Level	Protocols
Seasonal:	
April-August	• Restricted weekday usage of 20% of vehicles without catalytic converters • Implementation of a citywide traffic plan to minimize the effects of the vehicular restrictions
Episodic:	
Alert	• Restricted usage of 40% (weekdays) or 20% (weekends) of vehicles without catalytic converters • Prohibition on the use of uncertified residential wood or biomass heating units
Pre-Emergency	• Restricted usage of 60% (weekdays) or 40% (weekends) of vehicles without catalytic converters • Restricted usage of 20% (all days) of vehicles with catalytic converters • Require operational cessation of plants responsible for 30% of total stationary emissions of particulate matter • Physical Education classes and community sports activities may be suspended by the Ministry of Education • Implementation of more intensive traffic and public transportation plan • Increased enforcement of restrictions on mobile and stationary sources of air pollution • Increased and focused street sweeping and cleaning activities • Increased Metro service schedule implemented • Prohibition of the use of wood or other biomass for residential heating

Table 1.12: Episode Levels and Protocols in 1999 - Continued

Emergency	• Restricted usage of 80% (weekday) or 60% (weekend) of vehicles without catalytic converters • Restricted usage of 40% (all days) of vehicles with catalytic converters • Require operational cessation of stationary emissions sources (i.e., plants) responsible for 50% of total stationary emissions • Physical Education classes and community sports activities may be suspended by the Ministry of Education • Implementation of more intensive traffic and public transportation plan • Increased enforcement of vehicle usage restrictions • Increased and focused street sweeping and cleaning activities • Increased Metro service schedule implemented • Prohibition of the use of wood or other biomass for residential heating

Notes: Protocols were adjusted and updated periodically throughout the study period. Those described in this table were in place in 1999. This paper is primarily concerned with Pre-Emergency Episodes because these were both fairly common and involved the shutdown of stationary emissions sources (i.e., plants).

Chapter 2

Intertemporal Effects of Ambient Ozone Exposure on Human Performance

2.1 Introduction

Human health plays a significant role in determining both individual well-being and economic productivity. While the detrimental impacts of air pollution on human health have long been a subject of inquiry in the fields of public health, medicine, and economics (see for instance: Logan, 1953; Dimeo et al., 1981; Pope, 1989), studies of the effects of *in vivo* air pollutant exposures on human health have often been limited to the effects of single-period exposures, often on acute outcomes in at-risk populations.[1] While such studies clearly provide valuable information on the effects of air pollution exposure, little work has been done examining the effects of the ongoing exposures that more closely resemble the common exposure scenario of the general population. In order to better characterize the full effects associated with air pollution exposure in today's developed economies, a fuller investigation of cross-time, low-level ambient air pollution exposures on the healthy, working-age population is needed.[2] This project identifies sub-clinical effects of exposure to ambient *Ozone* on the physical productive capacity of young and fit adults (i.e.- college athletes) based on both contemporaneous exposure levels and average levels of exposure in the recent past, all in a multi-pollutant exposure framework.

In order to identify potentially small effects of low-level ambient pollution exposure on a healthy population, a precise and accurate outcome measure that can be compared across environmental conditions is required. Results

[1]Examples of work that has sought and found impacts of ambient air pollution exposure through the the tally of discrete events, such as hospital visits, asthma attacks, worker or student absences, or deaths include: Caiazzo et al. (2013); Chay et al. (2003); Currie and Neidell (2005); Neidell (2009); Schlenker and Walker (2012); Lleras-Muney (2010).

[2]There is a significant literature of lab work examining the exposure of (sometimes healthy) animal and human subjects in many different levels of many different pollutant. Several such studies that are particularly relevant to this project include: Devlin et al. (1991); Koren et al. (1989); Kehrl et al. (1987); Lippmann et al. (2005); McGrath (2000); Roger et al. (1985). Such studies will be addressed further, but the direct applicability of their results of small lab studies to uncontrolled environments and alternative outcome measures is unclear.

from outdoor, intercollegiate track & field meets provide the opportunity to compare closely measured performances of a healthy, working-age individual completing a highly uniform task under a variety of environmental conditions. Additionally, at the intercollegiate level, athletes can be thought of as experts in the tasks they complete, and these tasks are performed in a competitive setting. Together these facts suggest that measured outcomes closely reflect physical capacity, and thus any effect that is detected can be attributed directly to physical impacts of the air pollution exposure on the human body.

The large size of the track & field data source allows for the use of a multi-pollutant framework throughout the examination,[3] and the diverse range of track & field events tax the human physiology in distinct ways, providing the opportunity to examine heterogeneous effects of pollution exposure by task type. Additionally, knowing the location of a meet allows for the assignment of task-contemporaneous exposure levels, while the location of an athlete's home institution provides for the assignment of a measure of exposure levels during training. By estimating the effects of ambient pollution during training on contemporaneous ambient pollution effects I provide the first empirical evidence of physiological adaptation to ambient ozone.

The study of the contemporaneous effects of real-world exposure to ambient air pollutants on human performance - as a measure of the impacts on human well-being and a means of quantifying costs imposed on society - is a relatively recent development. Graff Zivin and Neidell (2012) is one of the first and only rigorous causal analyses to examine the effects of real-world ozone exposure on human performance, examining the productive output of agricultural workers on days with high levels of ambient ozone. Graff Zivin and Neidell find that a 10 parts per billion (*ppb*) decrease in ozone concentrations leads to a 5.5 percent increase in productivity among these workers. While data constraints prohibited analysis of other common pollutants, two

[3]Results of single pollutant analyses are easily confounded due to the close correlation of the ambient levels of some common pollutants, a fact highlighted in Currie and Neidell (2005).

other papers (Adhvaryu et al., 2014; Chang et al., 2014) have recently found economically significant impacts of indoor particulate matter on worker productivity.[4] Such results generally demonstrate real economic costs of ambient air pollution borne by society even when an acute health event is not observed.

In order to assess the impacts on human capacities of air pollution exposure at the low ambient levels common in developed nations today, this project begins with an examination of the impacts on human performance of five of the six criteria pollutants for which the Environmental Protection Agency (hereafter "EPA") is tasked with setting National Air Quality Standards.[5] The outcomes of interest to this analysis are competition results (hereafter: "event outcomes") of college-level Track & Field athletes competing under the umbrella of either the National Collegiate Athletic Association (NCAA) or the National Association of Intercollegiate Athletics (NAIA). Hundreds of collegiate meets[6] are held each year across the United States, yielding thousands of sanctioned event outcomes from athletes training in a diverse range of environmental conditions. The variation in the external environment of both the meets and athlete training locations, along with the uniform tasks and precisely measured outcomes, allows for detailed analysis focused on causal effects of environmental factors.

Using local measurements of ambient carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), particulate matter (measured as both PM_{10} and

[4]Contemporaneous exposure to air pollution has also been shown to negatively effect test scores (Lavy et al., 2012; Zweig et al., 2009), but no other work has investigated the impacts on human output or performance of *in situ* air pollution exposure.

[5]Lead is the sixth pollutant for which the EPA sets NAAQS, however since the phasing out of leaded gasoline in the United States, ambient lead has not been a significant health issueAgency (1994). As a result, monitor data on ambient lead levels are quite sparse, precluding the inclusion of lead in the broad analysis of this paper.

[6]I will adhere to the nomenclature common in track & field where an "event" is a specific type of competition - e.g.: shot put, long jump, 100 meter dash, etc. - and a "meet" is a gathering of different teams to compete, usually across multiple events. A round is a level of competition within an event (within a meet) in which all athletes who wish to move forward in the competition must participate - e.g.: qualifiers, finals, etc. - and a heat is an intraround grouping in which athletes compete for advancement (e.g.- one of two eight-person semi-final races to determine which athletes go on to compete in the final round).

$PM_{2.5}$), and sulfur dioxide (SO_2) from monitors across the United States, this project first examines the same-day effects of exposure to each pollutant on human performance. I find that contemporaneous ozone exposure negatively impacts performance in endurance events. These effects are statistically significant and robust to a diverse range of specifications, though they are quite small in magnitude. A fact that is likely due to the very low levels of ozone prevalent during the spring season in the United States when the majority of outdoor track & field meets are held.

Though small, the identifiable effects of ozone exposure on endurance events provide a basis for the examination of cross-time effects of ozone exposure. When training conditions are taken into account, in addition to contemporaneous pollution levels, I find an adaptation effect through which athletes training in higher-ozone environments are less impacted by ozone levels on competition days. The magnitude of this effect is such that athletes from the highest ozone locations are not negatively impacted by ozone levels at the average meet in the data, while athletes from home locations with lower average ozone levels suffer significant negative impacts from O_3 exposure on the average meet day in the data.

The broad goal of this project is to carefully characterize the effects of ambient ozone exposure (at levels common in the developed world) on human performance, then to examine how these effects are impacted by an individual's recent exposure history. Collegiate track & field provides a unique context within which to examine a tightly measured set of human performance metrics for which both current and recent-past ozone exposures can be assigned. The identification of significant adaptive effects of training under (moderately) higher ozone conditions suggests a much more nuanced model of the physiological effects of $Ozone$ exposure than is commonly considered and opens the door for much for further investigation of intertemporal effects of pollution exposure.

The remainder of this paper proceeds as follows. The next section presents background information on ozone, including physical sources, physiological channels of action in the human body, and potential negative impacts of exposure on humans at relatively low levels of exposure. The data sources and assignment of pollutant exposure are described in Section 3.3.1, and a categorical breakdown of track & field events is described in Section 2.3. Section 2.5 describes the empirical approach, and Section 2.6 lays out my findings as well as a number of robustness checks. The final section offers some discussion and concluding remarks.

2.2 Ozone and Human Well-Being

The majority of tropospheric (or low-level) ozone to which humans are potentially exposed is not directly emitted, but is instead formed in the atmosphere through a complex interaction of Volatile Organic Compounds, NO_x, and sunlight. Ozone is therefore commonly referred to as a secondary pollutant. A long atmospheric lifetime of nearly 22 days) leads to broad transport of O_3 from both its point of formation and the sources of its chemical components (Stevenson et al., 2006). The result of these dynamics is that peak O_3 concentrations are usually not directly attributable to local sources, but instead to baseline O_3 levels higher up in the atmospheric column. As such, the 8-hour averages observed at remote monitors - which are used for exposure assignment in this study - are considered to provide fairly accurate estimates of local outdoor exposure levels (Lippmann, 2009).

This paper follows in a venerable tradition of using athletic measures to examine health and performance effects of O_3 exposure, as the harmful effects of ambient O_3 were first identified in a study of high school track outcomes on high ozone days in California (Wayne et al., 1967).

Physiologically observable decrements in health are observable within a few hours of ozone exposure and may persist for hours or days after exposure, and a large proportion of O_3 that is inhaled appears to stay in the body

(Gerrity et al., 1988). Effects of exposure include reductions in lung capacity, increases in flow resistance of respiration, changes in epithelial permeability, and increased reactivity of bronchi (the passage which conducts air into the lungs) to other challenges (p. 870, Lippmann, 2009; Devlin et al., 1991; Kehrl et al., 1987; McDonnell et al., 1987). These reactions have been shown to impact lung and output performance among healthy adults undertaking a wide variety of physical activities including: exercise (Spektor et al., 1988), running (Selwyn et al., 1986), cycling (Brunekreef et al., 1994), hiking (Korrick et al., 1998), and agricultural work (Brauer et al., 1996; Graff Zivin and Neidell, 2012).

In addition to Wayne et al. (1967), which showed that track athletes were less likely to improve on previous performance levels when ozone concentrations were high, several studies have demonstrated reductions in athletic performance attributable to O_3 exposure. Schelegle and Adams (1986) showed that endurance athletes had a lower likelihood of completing a strenuous task when exposed to O_3 levels as low as 120 ppb, and demonstrated a significant decrease in the maximum exercise level of those exposed to 130 ppb O_3. Linder et al. (1988) also found that maximum exercise achieved by athletes decreased by 11% at an exposure level of 130 ppb O_3.[7] For comparison , I will be examining exposures to O_3 levels averaging approximately $35ppb$.[8] Generally, physical activity is considered to increase the level of absorption of ambient O_3 due to higher breathing volumes and increased flow rates of respiration (Giles and Koehle, 2014).

In vivo exposure to O_3 has been demonstrated to be more harmful than lab exposure, likely because of interactive effects with exposure to other pollutants (Spektor et al., 1988). This fact highlights the importance of carefully addressing potentially confounding factors when studying the effects of ozone

[7]For reference, the average O_3 levels match to event outcomes in this study is $44.58ppb$.

[8]This 35 ppb average reflects daily maximum levels of an 8-hour moving average of ozone concentrations, reflecting the measurement used by the relevant NAAQS, which is set such that the 4th highest report of this daily maximum measure should be below 75 ppb.

exposure outside the lab. While this examination focuses on the effects of $Ozone$, all analyses are done in a multi-pollutant exposure framework which explicitly takes into account five (of the six) categories of atmospheric pollutants which are regulated by the National Ambient Air Quality Standards (NAAQS): CO, NO_2, O_3, PM, and SO_2.[9] Such controls, along with flexible controls for temperature and humidity, and a wide range of fixed effects serve to increase the accuracy and precision of the estimated effects of ambient ozone exposure. For more discussion, see Section 2.5 below.

Finally, I am not the first to show an adaptation effect of multiple ozone exposures over time. There were a number of chamber studies conducted on very small samples (all had fewer than 30 participants) in the late 1970s and early 1980s that found that "repeated ozone exposure induces an adaptive response whereby subsequent ozone exposure induces little or no pulmonary function change" (Horvath et al., 1981, see also: Farrell et al., 1979). Interestingly there is some evidence that the threshold of exposure at which such adaption can occur is as low as between 20 and 40 ppb, a range of home ozone levels nicely covered by the data in this study(Dimeo et al., 1981). A single 8-person study looked at the effects of ambient (i.e.- non-chamber) exposures on adaptation, finding that Canadians were much more reactive to ozone that were Southern Californians (who presumably lived in higher day-to-day ozone conditions Hackney et al., 1977). This paper is thus the first to identify any sort of time-interactive ozone exposure effect on real-world outcomes, when both past and current exposures are to ambient levels rather than chamber doses.

[9] These pollutant, plus *lead*, are generally referred to as criteria pollutants. Lead is omitted because it is only sparsely monitored as it is no longer considered to be a significant threat at current levels (Agency, 1994).

2.3 Event Types

A diverse range of sporting events fall under the track & field moniker (known more generally as "athletics" outside the United States), with the only binding commonality being that each event is generally held on a 400m track or the field in the middle of such a track. For the purposes of this study, it is important to recognize that different events tax different aspects of human physiology. As such, I divide events into: sprint, endurance, and strength dominant categories based on the character of the effort required. Performance in sprint events requires intense effort over a short duration, and thus relies principally on fast twitch muscle fibers. Muscle fibers of this type are fueled primarily via anaerobic metabolization, which does not rely on the delivery of oxygen to fuel muscle function, and therefore should be unaffected by pollutants which impact respiration (Saltin and Gollnick, 2011). By contrast endurance events rely more heavily on slow twitch muscle fibers which (as the name suggests) contract more slowly, but are "well designed for prolonged activity", relying on aerobic metabolic processes for fuel (Saltin and Gollnick, 2011). Such aerobic pathways are likely to be impacted by decrements in oxygen delivery and cardiovascular performance, and thus we expect to see negative impacts on the performance of such activities from pollutants that act on these systems. Finally, strength events are primarily accomplished through a single exertion, the magnitude of which determines the caliber of the outcome. As such, strength events measure a maximum exertional capacity (setting aside technique and form factors) rather than an ability to draw upon and/or create energy reserves (Schulz and Curnow, 1988).

Practically, I will categorize "dashes" as sprint events, longer runs as endurance events, and throwing (javelin, shot put, discus, and hammer) and jumping (high jump, long jump, triple jump, and pole vault) events as strength events. The distinction between sprint and endurance events has been made between the 400 and 800 meter race lengths by naming conventions in college athletics (with the 400 meter race referred to as a "dash" and the 800 meter length called a "run"), but is generally borne out by research showing that athletes competing in the 400 meter distance draw approximately 60% of energy expenditures from anaerobic energy sources (and thus approximately 40% from aerobic energy sources) while competitors in the 800 meter event draw 60% of their energy from aerobic energy pathways (Duffield et al., 2005). The multi-sport events (heptathlon and decathlon) cannot be categorized in any of our three bins and are thus set aside for the current examination. See Table 2.1 for a complete categorization of the events examined by this paper.

Because the effects of ozone exposure are largely focused on the airways, and such effects appear to be exacerbated through increased rates and depths of breathing, it is natural to think that the detrimental effects of ozone exposure will be most pronounced in track & field events I have categorized as endurance events. The empirical analysis of this paper will thus focus on outcomes of such endurance events as the most likely to demonstrate identifiable effects of ozone exposure.

2.4 Data

It has long been recognized that results of sporting events provide a valuable source for precisely measured quantitative data on human performance under maximum effort(Hill, 1925). This paper relies on a newly constructed data set based on observational weather and air pollution data linked to more than 1.8 million event outcomes from 4,299 outdoor collegiate track & field meets held in the United States from January 2005 to June 2013.

The track & field event outcomes are provided by DirectAthletics, Inc. from the Track & Field Results Reporting System (TFRRS) database. The TFRRS is the official electronic reporting medium for NCAA track & field, and all official NCAA track & field results since mid-2009 have been posted to this system. Although no reporting mandate exists for NCAA results prior to

2009, or any NAIA results, the full set of event outcomes from many meets have long been posted to the database, leading to significant amounts of data from NAIA athletes and pre-2009 events. Table 2.2 lays out the number of event outcomes by year, as well as the count and characteristics of athletes captured in the data each year. It is important to note that teams (and thus athletes) are associated with specific leagues, but that a particular meet or event can (and often does) have athletes from different leagues competing against one another.

As the metrics for results reported in TFRRS differ between events, results for different events are not directly comparable. To provide a snapshot of the diversity of events captured in this data, Table 2.1 contains a list of event types for which more than 10,000 outcomes appear in the TFRRS data as well as the categorizations (as sprint, endurance, or strength) for such events which serve as the basis for the empirical analysis of this paper.

TFRRS captures:

1. Athlete data including: gender, year-of-eligibility, and location of home institution;

2. Heat-specific information including: round level, heat number, track type, altitude for venues over 3,000ft elevation, wind conditions, and athletes competing;

3. Event-outcome-specific data including: whether the event was finished, finishing place, and event outcome.

For each meet, contemporaneous weather and air pollution conditions are estimated based on monitor-level data from the National Oceanic and Atmospheric Administration (NOAA) and the Environmental Protection Agency (EPA) respectively. The weather data are taken from NOAA's Quality Controlled Local Climatological Data (QCLCD) service,[10] which provides daily

[10]NOAA QCLCD data are available here: http://cdo.ncdc.noaa.gov/qclcd

summaries of weather at approximately 1,600 stations across the United States. Weather conditions are assigned to each meet from the QCLCD monitor which is closest to the centroid of the meet's zip code. The mean distance from meet zip to matched QCLCD station in my sample is about 17 miles, and the farthest assigned station is 66.3 miles distant. Figure 2.1 shows the distribution of distances (in miles) from the centroid of the zip code in which an event outcome was recorded to the matched NOAA weather station. Table 2.3 reports summary statistics of the weather conditions assigned to track meets in the TFRRS data.

The number and distribution of active EPA monitors varies by pollutant and time, and thus meets (and athletes' home institutions) are matched to each pollutant individually. Daily data for CO, SO_2, NO_2, $Ozone$, PM_{10}, and $PM_{2.5}$(local conditions) were obtained from the EPA's AirData database.[11] Each meet day is matched (via the centroid of the zip code in which the meet was held) to the closest three active EPA monitors. Monitors more than 30 miles from the relevant zip centroid are dropped,[12] and valid reported values from the remaining station(s) are averaged using inverse distance weighting to estimate an ambient level for each meet/day and pollutant.[13] In order to capture the O_3 levels for athlete training grounds, the current year's average of 8-hour daily mean ozone levels is assigned from the closest active monitor (hereafter O_3^{Home}). The mean distances of matched monitors are summarized in Table 2.4. Table 2.5 provides summary statistics on the linked TFRRS/pollution data. The daily AirData data sets report the regulated daily statistic, which is different for each pollutant. Appendix Table 2.13 re-

[11] All EPA data were obtained at the daily summary level so that aggregations have already been made to the measure of each pollutant which is regulated by the NAAQS. Appendix Table 2.13 provides details of the relevant measures by pollutant. Data files are available from: http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html.

[12] The use of a different threshold distance does not significantly impact results.

[13] When multi-pollutant analyses are undertaken, all pollutants in the analysis must be matched for an event outcome to be included. This strong requirement explains the reduced number of observations used in each analysis as compared to those covered in the summary statistics tables.

ports the precise daily measures of each pollutant that are addressed by the respective NAAQS and are thus used in the analysis.

2.4.1 Cross-Event Comparability

In order to compare outcomes across events, a standardized result measure is calculated using the current world record for each event as a benchmark.[14] Specifically, I calculate a standardized performance measure of result i in event e using the following equation:

$$Std\_result_{ie} = \frac{|World\_Record_e - result_{ie}|}{SD(result_e)}$$

where $World\_Record_e$ is the current world record in event e and $SD(result_e)$ is the standard deviation of event e results reported in the TFRRS data. A decrease in this standardized measure means a move toward the world record, which is always an improvement. Men's and women's competitions are treated as separate events. Outlier results are removed from the analyzed sample by three times dropping observations with a standardized result ¿5 and recalculating the standardized result.

2.4.2 Special Considerations

A major shortcoming of the TFRRS data are that the exact day on which an event outcome took place is not recorded. Instead, each result is mapped only to a meet, for which the start and end date are recorded. Thus, pollution and weather levels are assigned to a meet, using the mean levels from all days over which the meet took place (except for wind assist which is captured in TFRRS data for each heat).

This limitation has the potential to introduce significant measurement error into the analysis, and while a small majority of outcomes in the data

[14]World records were current as of June 2014, as reported by the International Association of Athletics Federations (IAAF) on its website: http://www.iaaf.org/records/by-category/world-records.

(just over 50%) are from meets that took place on a single day, and a full 83.3% of recorded outcomes are from meets of length 1 or 2 days, the rough assignment of pollution and weather conditions to event outcomes should be kept in mind throughout this analysis. See Figure 2.2 for a breakdown of event outcomes by meet length.

A number of distinct advantages of the TFRRS data are also worth highlighting. First and foremost, the data captures exacting outcome measurements of a uniform task repeated by the same individual multiple times under different external conditions. Such a setup makes intuitive the idea that differences in outcomes may be attributable to differences in external conditions.

Secondly, the elite level of the athletes whose performance is captured in the TFRRS data affords a much higher level of individual consistency from day-to-day than we would expect in many other settings. Such consistency ensures that more of the observed between-meet variation in performance might be attributable to observable external factors (rather than random variation in athlete performance).

Third, competitors in collegiate level track & field events are likely performing at, or very near, the maximum thresholds of their capabilities. Thus, even small impacts on physical capacity might be identifiable through the examination of event outcomes in the data. Additionally, the highly competitive and high-stakes nature of the studied outcomes reduces the likelihood of significant behavioral confounders such as shirking (a major problem in many studies which seek to analyze impacts on human output). Impacts of the external environment on performance would be more easily missed in the examination of other, less demanding or less competitive, types of activities.

Finally, the use of track & field event outcomes allows environmental conditions at each athlete's home institution to be mapped to each event outcome as well as contemporaneous conditions. This allows for the examination of intertemporal effects of ozone exposure using both levels during training as well as at competitions.

2.5 Empirical Methods

This investigation will proceed in two stages. First, a straightforward regression of event outcomes on pollution levels of our six pollutant measures will be estimated. Second, an analysis of adaption or exacerbation effects of training conditions will be undertaken to investigate whether ozone exposure during training might modulate the effects of task-contemporaneous ozone exposure.

The core of this empirical investigation rests on two implementational strengths of the setting and data. First the large size of the TFRRS data set allows for all analyses to be conducted in a multi-pollutant framework. Thus, even when certain pollutants will not be of central note in the discussion going forward, the levels of all six pollutant measures enter each regression to serve as controls. This reduces the potential for mis-attribution of the effects of one criteria pollutant to another correlated criteria pollutant.

Second, this analysis will rely on a number of fixed effects to account for unobserved heterogeneity across a diverse range of outcome characteristics. The comparison of event outcomes between different individuals introduces a host of unobservable athlete-specific characteristics that might explain the observed variation in outcomes. By using an athlete fixed effects approach, time-constant athlete-specific characteristics are removed from consideration. Because athlete performance likely changes over the years (experience) and through the season (practice), I will also include year-of-eligibility and week-of-season fixed effects. Venue fixed effects are included to control for altitude, track surfaces, and other time-constant factors that might effect outcomes. Finally, because analyses are conducted across events, event fixed effects are included to absorb any additional variation that exists in the standardized outcome variable used for analysis.

2.5.1 Econometric Specifications

To focus the analysis on ambient air pollution levels, a range of other external factors that differ between meets and heats are also controlled for: temperature, humidity, and wind conditions. The following specification is estimated on separate samples for each event type and gender. Letting i represent the athlete in year $z \in [1, 2, 3, 4]$ of athletic eligibility, in the hth heat, of the eth event, at the mth meet, at the vth venue, in week w of season t; the central empirical specification is as follows:

$$Std\_rslt_{ihemvwt} = \alpha_i + \sigma_{it} + \kappa_w + \nu_v + \rho_e + \mathbf{PC}'_{\mathbf{m}}\hat{\beta} + \mathbf{X}'_{\mathbf{m}}\hat{\delta} + \lambda * W_{hem} + \varepsilon$$

where the standardized result is regressed on a vector of linear ambient pollution levels, $\mathbf{PC}_m$, as well as a number of controls. In order to remove the effects of other, non-pollutant environmental factors, wind-assist (W_{hem}: measured for each heat of each event) and indicators for 5-degree bins of average temperature and 2.5 degree bins of dew point temperature (both in $\mathbf{X_m}$) are included in all regressions.[15] Fixed effects are included for each athlete (α_i); year-of-eligibility (σ_{it}); week-of-season (κ_w); venue (ν_v); and event/gender (ρ_e). ε is the outcome specific error term. Standard errors for all analyses are two-way clustered at the meet and athlete levels. The lack of within-meet variation in the pollution levels requires clustering at the meet level, while the possibility of serial correlation in athlete performance necessitates clustering by athlete (Cameron et al., 2011).

[15]Given that ozone is generated through a chemical reaction catalyzed by sunlight, and temperature has direct (and likely non-linearly, see:Hancock, 1989) effects on athlete performance, it is particularly important that temperature is controlled for flexibly. Thus the bin-indicator approach is central to the identification strategy as it minimizes the functional form assumptions neccesary for analysis. Other specifications of temperature and humidity controls are considered as robustness checks but do not markedly impact estiamtes.

2.5.2 Intertemporal Exposure

The main specification has followed the standard approach of analyzing the effects of a single time-frame of pollution exposure. In this investigation the time-frame of interest has been contemporaneous to the task completion, though other papers look at the effects of exposures during certain stages of development on outcomes later in life (see Currie et al. (2014) for a number of such examples). I turn now to the task of assessing if and how ambient ozone exposure during training may modulate the effects of exposure contemporaneous to competition.

To start, the annual mean ambient O_3 level at each athlete's home institution - near where most training likely takes place - is interacted with the day-of-meet O_3 level for each event-outcome. This interacted specification simply adds the term: $O_3^{home} * O_3^{meet}$ to the main specification laid out above. All other pollutant levels (including O_3^{meet}) still enter as before, though O_3^{home} is not added directly as the variation of this variable is largely absorbed by the athlete and year-of-eligibility fixed effects.

As a less-parametric approach to examining how exposure to O_3 during training might modulate the effects of day-of-competition O_3 exposure, I examine how the estimated effects of O_3^{meet} levels differ for athletes by home-ozone level. In particular, athletes are divided into 10 groups based on O_3^{home} levels falling into 2.5 ppb intervals from 20 ppb to 40 ppb. Indicators for these ten bins are then interacted with the meet-O_3 levels to compare how an increase in O_3 on a meet day might differentially impact an athlete that trains at an average O_3level of 24 ppb compared to an athlete that trains at an average of 32 ppb. All other controls and fixed effects remain the same as in the main econometric specification.

2.6 Results

2.6.1 Contemporaneous Results

Table 2.6 presents the coefficient estimates of the multi-pollutant model run for each gender and event type (i.e.- endurance, sprint, and strength). These estimates validate my focus on the effects of O_3 exposure on outcomes in endurance events. For men, we see detrimental effects of 0.011 standard deviation in outcomes across endurance events from a 10 ppb increase in ozone, and for women we see a 0.0095 standard deviation effect of the same change in ozone. Both estimates are significant at the 5% level, but are quite small.

For perspective, I am working with standardized results, so the standard deviation within events equals 1, and the mean of the standardized result outcome for male endurance events is 2.74. Thus, these results suggest that a 10ppb increase in ozone causes a 0.40% change in outcome. Given that the difference between first and second place finishers in the data is an average of 6.1% of the mean outcome, the male effect of ozone exposure estimated here is about 6.6% of the average difference between the first and second place finishers in men's endurance events.

It is worth mentioning that we also see mostly detrimental effects of exposure to particulate matter. While the direction of these results fits with existing literature (for example: Gold et al. 2000; Holguín et al. 2003; Ghio et al. 2000; Lippmann et al. 2005), the small magnitude and inconsistent significance suggest that we don't quite have the statistical power to really examine the effects of PM in this context.[16]

The estimates of the effects of both CO and SO_2 are in the direction we might expect for pollutants labeled by the EPA as harmful (i.e.- they negatively impact performance), but all estimates are extremely small in magnitude and

[16]Mean levels of PM_{10} and $PM_{2.5}$ Local Conditions are 19.29 and 8.47 $\mu g/m^3$ respectively, well below the NAAQS levels for these measures of 150 and 35 $\mu g/m^3$. See Appendix Table 2.13 for a more thorough presentation of the details of the NAAQS.

none are statistically significant. The CO results are unsurprising given that remote monitor measurements of CO are generally thought to be poor proxies for exposure even at locations that are quite nearby, which adds even more noise to the estimation procedure and raising the size of effect that would be detectable (Lippmann, 2009). In this investigation, SO_2 is only considered as part of a multi-pollutant exposure model, and negative effects of ambient SO_2 (when detected at all) tend to disappear when particulate matter levels are added into consideration within a multi-pollutant framework (Lippmann, 2009). Thus, again, it is unsurprising that my estimates do not identify effects from SO_2.

More unique to this investigation are the consistently significant beneficial effects of NO_2 exposure across genders and event types. These effects are small in magnitude, but their consistency merits consideration. While this is not the first study to find beneficial effects of NO_2, (Linn et al., 1985) found reduced blood pressure during exercise associated with exposure to NO_2, most studies find no significant effect of NO_2exposures to concentrations typical of ambient levels on healthy adults (Bascom et al., 1996; Hesterberg et al., 2009; Langrish et al., 2010). It is likely therefore, that NO_2 levels in this setting are serving as a proxy for ambient NO, which has been shown to function as a selective vasodilator in the lungs following inhalation, reducing arterial pressure in the lungs and increases blood oxygenation (Rossaint et al., 1993, 2014; Griffiths and Evans, 2005). Inhaled NO, (indicated as iNO) has been approved for use in treating infants with a number of pulmonary issues resulting in low respiratory function (e.g.-Kinsella et al., 2006; Roberts et al., 1997), and many studies have found ameliorative effects of iNO on adults suffering from hypoxemia (low levels of oxygen in the blood) and other issues with lung function (see for example:Rossaint et al., 2014; Gómez et al., 2013; Teman et al., 2015). While the data used for this study does not provide measurements of NO levels, a previous iteration of this work relied on a separate EPA data source (EPA AQS Data) in which the correlation between NO and NO_2 at studied track meets was shown to be about: 0.7.

It is important to note that there is not a high degree of correlation between ozone levels and other pollutants in the data. Table 2.7 reports the correlations between matched pollutant levels in the data, and we see that all correlations are very near zero.

Those familiar with pollutant correlation levels will notice immediately that Table 2.7 does not reflect typical pollutant concentration relationships (see for example: Levy et al., 2014). This is due to the fairly narrow time frame in which outdoor collegiate track meets are held each year. As Table 2.3 reports, nearly 98% of outdoor track meets are held in March, April, or May. Spring does not tend to be the high risk season for any major pollutants in North America, and thus it is rarely the focus (or even included) in air pollution studies (for example, Lavy et al., 2012 takes samples in all three other seasons to examine pollutant correlation relationships). The unique pollutant correlations during the time of this study allow for a additional confidence in the multi-pollutant results, since high levels of correlation amongst pollutants can make causal assignment problematic during other times of the year.

The low correlation is particularly important in the current context for for NO_2 given the significant effects identified for exposure that move in the opposite direction of those for ozone. If these two were closely correlated, it could be that the estimation was being confounded by random differences between two otherwise similar variables. To ensure that the multi-pollutant estimation framework is not inappropriately driving results, I re-estimate the effects of ozone on endurance while excluding various groups of the other pollutant controls. The results of this exercise are presented in Table 2.8 and show that the ozone effects for endurance events are robust to the inclusion or exclusion of various pollutant controls, with the estimated coefficients on ozone varying very little between models. Though the inclusion of NO_2 moderates the estimated effects of ozone, this demonstrates that NO_2 is not driving the results of interest to this investigation and further highlights the importance of controlling for other pollutants in my analyses.

Another possible fragility of models seeking to identify the effects of ozone on humans, is the close relationship between ozone and temperature and human performance and temperature. The main specification relies on flexible controls for temperature, but to ensure that it is not these controls which are driving the reported results, the main specification is re-estimated allowing average temperature to enter as a quadratic or alternatively allowing both daily maximum and minimum temperatures to each enter as a quadratic. Additionally, rather than relying on binned-level-indicators of dew point to control for humidity, I add in indicators for heat risk which provides four categories of weather conditions which have been identified as imposing Low, Moderate, High, and Extreme risk of exertional heat illness based wet-bulb globe temperature (from the NOAA data), a composite metric which factors in temperature, humidity, wind, and radiation effects on humans(Binkley et al., 2002). Each specification is run both with and without wind assist and travel-distance controls, and the results are reported in Table 2.9. Again we see that the ozone effects estimated in the main analysis are quite robust to the details of the specification.

Finally, I present the results of the main specification on each individual endurance event in Table 2.10. These estimates show that the main effects exist even in the raw result data, and are not somehow a residual of the standardization process. Generally, we see that larger and more significant effects are found for longer events. This fits nicely with the reasoning which has driven our focus on endurance events over sprint and strength events.

2.6.2 Intertemporal Results

Turning to the question of how training conditions modulate the effects of contemporaneous-ozone exposure, I look first at the interaction of home average exposure levels with contemporaneous O_3 exposure. Although the baseline effects of ozone exposure detected in this paper are small in magnitude, their significance and robustness serve as a solid basis for examining intertemporal relationships between current and past exposure levels. Table 2.11 reports the regression coefficients on O_3^{Meet} and the $O_3^{Meet} \times O_3^{Home}$ interaction for men and women's endurance events. Importantly we see the same magnitude of detrimental impacts of O_3^{Meet} at mean O_3^{Home} that we saw in the main specification.[17] The (relatively) large and significant negative coefficients on the interaction term tell a story of increased home-ozone mitigating the negative effects of day-of-meet-O_3 levels. These results are significant for both men and women, large compared to the main effect of O_3^{Meet}, and novel in the literature. This relationships suggests a sort of acclimatization or adaptation effect of living/training with regular ozone exposure.

Because most athletes tend to compete in meets that are fairly close to their home institution, it may be that meet-ozone is highly correlated or well predicted by variation O_3^{Home}. If such were the case, the usefulness of the above results would be questionable. However, in the data, there is only a correlation coefficient of 0.116 between meet and home ozone levels, and a regression of meet-ozone on home-ozone levels has an r-squared of only 0.0136, suggesting that home-ozone exposure is not in fact predictive of meet-ozone levels.

To further investigate the nature of the relationship between home and meet ozone, we examine the estimated effects of O_3^{Meet} on endurance results separately for groups of athletes with different home ozone levels. Specifically, athletes are divided into groups based on 2.5 ppb bins of average home ozone levels. Table 2.12 reports the effects of contemporaneous ozone separately for the 10 groupings based on home ozone levels. Athletes in the first bin are from home institutions with an average annual ozone level for the current year of ¡20 ppb. These are the athletes training in the cleanest home environs (at

[17] Assessed at mean levels of O_3^{Home}, 28.72 ppb, the effect of O_3^{Meet} is 0.0434 + 2.872 * -0.0117=0.00980 for men and 0.0406 + 2.872*-0.0112=0.00843 for women.

least as far as ozone levels), yet we see distinctly that men for such places have competition outcomes that are most harmfully affected by ozone on meet days. The estimates for both men and women suggest that athletes training in lower-ozone locales face stronger negative impacts of ozone exposure during endurance competition. In fact, the analysis suggests that athletes training in the lowest-ozone home institution face detrimental impacts of ozone, while those training in ozone levels comparable to the mean ozone at meets (around 35 ppb) do not face statistically significant effects of ozone exposure during competition of endurance tasks.

Given the relatively low ozone levels that prevail in the United States today, there is very little room in this study to assess if and when such an adaption effect might be swamped by long term damage done by regular exercise in a highly polluted environment.[18] That being said, it appears that the falling impacts of day-of-meet ozone may begin turning around with the 35-37.5 ppb bin for both men and women. Figure shows the bin-by-bin plots of ozone effects for Power and Strength events in addition to Endurance events. This graph provides some suggestive evidence of a change in the the adaption effects of ozone exposure during training around the 35 ppb mark.

2.7 Discussion and Conclusion

This paper has broadly investigated the effects of exposure to low-level ambient air pollution on human performance across a diverse range of physically taxing tasks. Due to the physically demanding and competitive nature of the tasks analyzed, performance can be thought of as closely reflecting the health and functionality of certain physiological systems. I find that the current ambient levels of CO, SO_2, and particulate matter that exist across the United States in the springtime do not have meaningfully negative impacts

[18]There definitely exists evidence for negative health effects of long term exposure to elevated O_3 levels. See for example: Jerrett et al., 2009)

on human health or performance of even the most demanding tasks among collegiate athletes. As noted, the beneficial results identified for NO_2 indicate that it is likely serving as a proxy for levels of NO which has a known ameliorative effect on the pulmonary system when inhaled. Finally, I show robust and significant detrimental effects of contemporaneous exposure to ozone on performance of endurance events. Given the low pollution levels that prevail in this study, and the health and fitness levels of the studied population, it is unsurprising that the magnitude of the identified results is fairly small and localized in endurance events which rely most heavily on the aerobic metabolic pathways which ozone exposure affects.

This is the first examination of the interactive effects of in situ pollutant exposures over disparate time-frames. Using the significant effects of contemporaneous ozone exposure on performance in endurance events as a basis, I am able to examine how these effects are impacted by ozone exposure during training. I find convincing evidence of differential impacts of contemporaneous ozone exposure across mean ozone levels in the home environment. Specifically I show that individuals that train in the lowest-ozone environments are most sensitive to the effects of ozone exposure on competition days, suggesting a sort of adaptation effect of training in the presence of ozone. Just as training at altitude reduces the deleterious effects of performance at altitude (Geiser et al., 2001), it appears that training with ambient ozone may mitigate the negative impacts of ozone on performance.

The fact that I find an apparent adaptation effect - whereby athletes that train in higher ozone environments are less affected by ozone at competitions - raises a number of questions about the regulation and costs of exposure to ambient ozone. Could reducing average levels of ambient ozone increase the costs associated with a spike in ozone levels? Is compliance with a threshold a useful way to regulate ozone if constant levels of ozone are less harmful than variable levels? How are the long term detrimental impacts of ozone exposure related to the medium term interactions shown here? At what levels of average exposure are the adaptive effects of ozone exposure trumped by damage inflicted? Should travelers planning on visiting a high ozone environment attempt to acclimatize beforehand?

There is a small amount of evidence suggesting that adaptive effects of ozone exposure are short lived - on the order of 2 weeks (Horvath et al., 1981) - which rules out the feasibility of any sort of ozone-inoculation because of the frequency which such treatments would have to take.

Most importantly, it is no longer clear for developed countries that have already achieved fairly low ambient average ozone levels, whether further reductions are beneficial on the whole to health. This is a first order issue which goes beyond cost-effectiveness investigations, as it could be that further levels of ozone reductions could have net-negative impacts on health in addition to the economic costs of undertaking such cleanup. Thus, better studies of the long-term effects of ambient ozone exposure throughout and across stages in life are needed to fully characterize the effects of ongoing ambient exposures on current and future outcomes.

2.8 Acknowledgements

Chapter 2, in full, is currently being prepared for submission for publication of the material. Jamie Mullins, the dissertation author was the primary investigator and author of this material.

2.9 Tables and Figures

Table 2.1: Event Types with Greater than Ten Thousand Outcomes

Type	Event	Female Results	Male Results	Total
Sprint Events	100m Dash	68,147	76,578	144,725
	100m Hurdles	59,612	0	59,612
	110m Hurdles	0	49,041	49,041
	200m Dash	83,191	72,495	155,686
	400m Dash	54,964	67,530	122,494
	400m Hurdles	40,574	41,747	82,321
Endurance Events	800m Run	79,077	78,538	157,615
	1,500m Run	74,352	82,967	157,319
	3,000m Run	12,716	9,676	22,392
	5,000m Run	42,814	50,255	93,069
	10,000m Run	12,155	15,355	27,510
	3,000m Steeplechase	20,695	25,657	46,352
Strength Events	Discus Throw	51,053	61,086	112,139
	Hammer Throw	47,097	44,532	91,629
	Javelin	50,661	48,824	99,485
	Shot Put	58,075	56,029	114,104
	High Jump	39,198	31,897	71,095
	Long Jump	54,921	51,262	106,183
	Pole Vault	29,869	32,585	62,454
	Triple Jump	31,805	26,872	58,677
	Total	910,976	922,926	1,833,902

Notes: Table only includes event types with more than 10,000 event outcomes captured in the TFRRS database from January 2005-June 2013 and matched to pollution and weather monitor days. Categorization of events is discussed further in Section 2.3. Relay (i.e.-multi-athlete) and indoor events are omitted from the analysis and therefore not reported in the table above.

Table 2.2: Athletes by Year

Year	Event Outcomes	Athlete Count	Gender		Year of Eligibility				NAIA	NCAA	NCAA	NCAA
			M	F	FR	SO	JR	SR		Div. I	Div. II	Div. III
2005	22,035	10,306	5,356	4,950	3,459	2,553	2,208	2,086	290	6,765	1,259	1,992
2006	39,792	16,844	8,519	8,325	5,610	4,462	3,637	3,132	551	9,752	2,554	3,987
2007	49,029	19,358	9,928	9,430	6,542	4,970	4,222	3,624	898	10,076	3,293	5,091
2008	63,917	22,120	11,419	10,701	7,737	5,873	4,524	3,985	1,312	10,771	3,732	6,305
2009	99,617	28,946	15,067	13,879	9,576	7,820	6,221	5,329	1,687	14,347	4,578	8,334
2010	336,475	44,968	23,618	21,350	16,252	11,769	9,463	7,484	3,663	18,616	8,552	14,137
2011	392,948	48,340	25,107	23,233	16,861	12,727	10,474	8,277	5,014	18,807	9,143	15,376
2012	397,789	49,370	25,311	24,059	17,192	13,103	10,732	8,340	5,105	19,150	9,329	15,786
2013	399,005	50,652	25,742	24,910	17,826	13,060	10,872	8,891	5,646	19,014	9,517	16,475

Notes: Counts include outdoor results matched to pollution and weather stations.

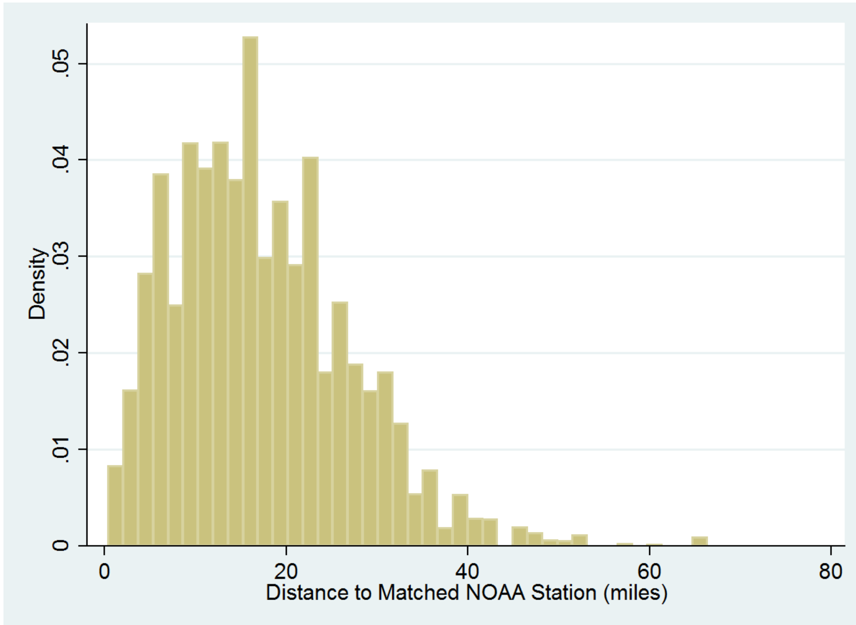


Figure 2.1: Distribution of Distance to Matched Weather Data Source by Event Outcome

Notes: Chart includes all outdoor meets which could be linked to a weather monitor. Only the closest NOAA station is linked to each meet.

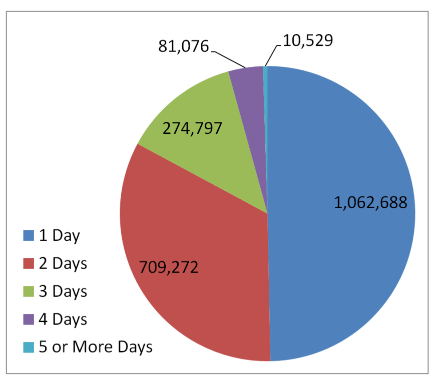


Figure 2.2: Event Outcomes by Meet Length in Days

Notes: Chart includes all outdoor events.

Table 2.3: Yearly Summary Stats Including Weather

Year	Meets	Athletes	Total Event Outcomes	Max Temp	Min Temp	Mean Temp	Dew Point	Meets by Month						
								Jan	Feb	Mar	Apr	May	Jun	Dec
2005	61	10,531	22,410	65.99	43.62	55.04	38.77	0	4	14	35	8	0	0
2006	117	17,840	45,069	67.42	45.34	56.65	40.45	2	0	30	58	26	1	0
2007	150	21,438	57,756	65.56	42.71	54.37	36.13	4	9	41	64	32	0	0
2008	183	25,128	75,648	67.06	44.14	55.82	38.65	4	5	51	86	35	0	2
2009	260	30,573	109,603	66.90	44.40	55.91	40.07	2	6	63	135	53	1	0
2010	840	45,027	340,826	68.00	44.38	56.43	40.60	1	16	183	463	173	4	0
2011	878	48,350	394,689	64.73	43.18	54.18	40.94	0	11	185	513	165	4	0
2012	925	49,368	398,795	68.94	46.44	57.98	43.84	0	11	265	479	166	4	0
2013	885	50,659	402,235	63.96	40.45	52.45	36.86	0	9	249	467	158	2	0
Total	4,299	298,914	1,847,031	66.51	43.85	55.42	39.59	13	71	1081	2300	816	16	2

Notes: Counts include events at outdoor meets only. Weather values are averages across all outdoor event outcomes which could be matched to a NOAA station.

Table 2.4: Mean Distance to EPA Monitors Used in Calculating Pollutant Levels (in Miles)

	Closest Monitor	2nd Closest Monitor	3rd Closest Monitor
CO	9.39	12.69	15.50
SO_2	10.87	15.62	17.39
NO_2	9.33	13.47	14.71
O_3^{Meet}	8.00	12.53	15.79
$PM_{2.5}$	7.52	12.03	14.66
O_3^{Home}	15.24		

Notes: Distances are measured in miles. For each pollutant, the first reported value is the distance from the centroid of the zip code in which a meet took place to the closest monitor which was active at the time. Similarly for 2nd and 3rd closest monitor distances. Home institution O_3 distances are higher because no matches are dropped for distances over some threshold. Additionally, many institutions are large enough to field teams but don't have facilities to host meets. Such institutions tend to be in more rural areas, further from pollution monitors.

Table 2.5: Pollutant Levels: Counts of Matched Event Outcomes & Mean Values

	Total Event Outcomes	**Reported Statistics**	**CO** (*ppb*)	$\mathbf{SO_2}$ (*ppb*)	$\mathbf{NO_2}$ (*ppb*)	$\mathbf{O_3^{Meet}}$ (*ppb*)	$\mathbf{PM_{10}}$ ($\mu g/m^3$)	$\mathbf{PM_{2.5}}$ ($\mu g/m^3$)	$\mathbf{O_3^{Home}}$ (*ppb*)
2005	22,410	Matched Outcomes	16,495	15,891	15,411	20,279	19,013	18,707	22,351
		Mean Level	496.44	2.66	14.59	34.36	19.03	8.41	28.06
2006	45,069	Matched Outcomes	25,143	23,128	23,133	33,701	29,909	33,491	44,926
		Mean Level	420.27	2.43	12.12	33.69	19.95	9.52	28.57
2007	57,756	Matched Outcomes	31,008	31,161	28,114	41,814	36,801	41,971	57,706
		Mean Level	395.34	3.20	14.06	33.08	21.86	9.40	29.15
2008	75,648	Matched Outcomes	42,963	41,191	38,843	53,375	46,989	52,419	75,467
		Mean Level	425.18	2.62	12.54	36.93	22.01	10.61	29.03
2009	109,603	Matched Outcomes	63,325	62,263	55,217	83,535	74,725	83,468	109,365
		Mean Level	371.47	2.19	12.11	35.48	19.31	9.01	27.16
2010	340,826	Matched Outcomes	217,389	218,393	197,568	289,067	235,308	283,820	340,499
		Mean Level	356.16	1.40	10.04	35.48	20.30	8.68	29.36
2011	394,689	Matched Outcomes	250,718	254,561	221,767	339,206	265,947	318,992	394,258
		Mean Level	316.07	1.22	9.54	33.77	18.19	8.43	29.03
2012	398,795	Matched Outcomes	238,501	250,844	218,286	336,640	263,401	317,658	398,497
		Mean Level	317.13	1.00	9.44	33.58	19.15	8.11	29.77
2013	402,235	Matched Outcomes	230,406	271,776	229,427	343,690	260,922	320,014	401,817
		Mean Level	305.85	0.94	9.21	35.01	18.71	7.95	28.37
Total	1,847,031	Matched Outcomes	**1,115,948**	**1,169,208**	**1,027,766**	**1,541,307**	**1,233,015**	**1,470,540**	**1,844,886**
		Mean Level	**378.21**	**1.96**	**11.52**	**34.60**	**19.83**	**8.90**	**28.72**

Notes: Pollutant counts represent the number of event outcomes for which a valid measure was matched. Pollutant counts are less than the total number of event outcomes because of seasonally (or temporarily) inactive monitors and meets which are held further than 30 miles from the nearest active monitor for a given pollutant. The second reported number for each pollutant/year is the average concentration in the data. With the exception of O_3^{Home}, all levels are the average across the duration of a meet. O_3^{Home} is the average annual level of ozone at the home institution of the athlete.

Table 2.6: Multi- Pollutant Regressions by Event-Type

	Endurance		**Power**		**Strength**	
	Male	Female	Male	Female	Male	Female
CO	**0.0000548**	**0.0000177**	**0.00000319**	**0.0000434**	**0.0000164**	**0.0000378**
(*ppb*)	(0.000041)	(0.000035)	(0.000037)	(0.000034)	(0.000030)	(0.000029)
SO_2	**0.000000138**	**0.00000006**	**-0.00000179**	**-0.000000209**	**0.00000164**	**0.000000231**
(*ppb*)	(0.000003)	(0.000003)	(0.000003)	(0.000003)	(0.000002)	(0.000002)
NO_2	**-0.00469*** **	**-0.00320*** **	**-0.00474*** **	**-0.00492*** **	**-0.00257*** **	**-0.00173*** **
(*ppb*)	(0.000881)	(0.000813)	(0.000784)	(0.000702)	(0.000644)	(0.000633)
O_3	**0.0107** **	**0.00946** **	**-0.00141**	**-0.00266**	**-0.00662* **	**-0.00212**
(10 *ppb*)	(0.004790)	(0.004470)	(0.004470)	(0.003810)	(0.003390)	(0.003640)
PM_{10}	**0.000790** **	**0.000498**	**0.000626* **	**0.000454**	**0.00045**	**0.000521** **
($\mu g/m^3$)	(0.000381)	(0.000348)	(0.000334)	(0.000297)	(0.000286)	(0.000262)
$PM_{2.5}$	**-0.000307**	**0.000253**	**-0.000641**	**0.000642**	**0.0000403**	**0.000442**
($\mu g/m^3$)	(0.000974)	(0.000935)	(0.000935)	(0.000787)	(0.000786)	(0.000768)
Observations	104,564	98,298	122,929	127,701	140,216	147,455
Athlete #	18,618	17,311	18,195	17,582	17,210	17,606
Adjusted r^2	0.861	0.887	0.802	0.879	0.81	0.819

Notes: Robust standard errors in parenthesis are two-way clustered at the meet and athlete levels. All regressions include controls for wind assist, travel distance, temperature ($5^\diamond F$ bin indicators), dew point ($2.5^\diamond F$ bin indicators), home field advantage, experience (year-of-eligibility fixed effects), practice and meet importance (week-of-season fixed effects), venue characteristics (venue fixed effects), gender and event-specific peculiarities of the normalized outcome measure (event fixed effects), and unobserved athlete heterogeneity (athlete fixed effects). Measures of ambient pollution are averages of ambient levels across meet days.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.7: Correlation Matrix of Air Pollution Levels

Correlations:	CO	SO_2	NO_2	$Ozone$	PM_{10}	$PM_{2.5}$
CO	1.0000					
SO_2	0.1912	1.0000				
NO_2	0.4483	0.3224	1.0000			
$Ozone$	-0.1572	0.0479	-0.2051	1.0000		
PM_{10}	0.1859	0.0768	0.3517	0.0343	1.0000	
$PM_{2.5}$	0.2137	0.1998	0.3618	-0.0224	0.4703	1.0000

Notes: Correlations of air pollution concentrations across matched event outcomes.

Table 2.8: Multi- Pollutant Robustness

	Endurance Events		Endurance Events		Endurance Events		Endurance Events		Endurance Events		Endurance Events	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Ozone	0.0200***	0.0154***	0.00921**	0.00879**	0.0181***	0.0136***	0.00951**	0.00861**	0.0107**	0.00976**	0.0107**	0.00946**
(10 *ppb*)	(0.00343)	(0.00295)	(0.00419)	(0.00364)	(0.00422)	(0.00362)	(0.00466)	(0.00416)	(0.00484)	(0.00434)	(0.00479)	(0.00447)
$\mathbf{NO_2}$			-0.00422***	-0.00313***			-0.00431***	-0.00304***	-0.00473***	-0.00318***	-0.00469***	-0.00320***
(*ppb*)			(0.00065)	(0.00058)			(0.00074)	(0.00067)	(0.00083)	(0.00076)	(0.00088)	(0.00081)
$\mathbf{PM_{10}}$					0.00036	0.000507*	0.000801**	0.000584*	0.000695*	0.00045	0.000790**	0.000498
($\mu g/m^3$)					(0.00034)	(0.00030)	(0.00036)	(0.00033)	(0.00036)	(0.00034)	(0.00038)	(0.00035)
$\mathbf{PM_{2.5}}$					-0.00259***	-0.00221***	-0.000771	-0.000177	-0.000567	0.000102	-0.000307	0.000253
($\mu g/m^3$)					(0.00083)	(0.00076)	(0.00093)	(0.00087)	(0.00097)	(0.00090)	(0.00097)	(0.00094)
CO									0.0000576	0.0000209	0.0000548	0.0000177
(*ppb*)									(0.00004)	(0.00003)	(0.00004)	(0.00004)
$\mathbf{SO_2}$											0.000000138	5.68E-08
(*ppb*)											(0.000003)	(0.000003)
Observations	202,775	187,239	131,482	122,276	141,692	131,464	113,940	106,996	108,880	102,349	104,564	98,298
Adjusted r^2	0.858	0.885	0.858	0.886	0.859	0.887	0.86	0.886	0.861	0.888	0.861	0.887
# of Athletes	27,056	25,348	21,290	19,798	22,511	20,953	19,524	18,152	18,983	17,628	18,618	17,311

Notes: Robust standard errors in parenthesis are two-way clustered at the meet and athlete levels. All regressions include controls for wind assist, travel distance, temperature ($5^\circ F$ bin indicators), dew point ($2.5^\circ F$ bin indicators), home field advantage, experience (year-of-eligibility fixed effects), practice and meet importance (week-of-season fixed effects), venue characteristics (venue fixed effects), gender and event-specific peculiarities of the normalized outcome measure (event fixed effects), and unobserved athlete heterogeneity (athlete fixed effects). Measures of ambient pollution are averages of ambient levels across meet days.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Robustness to Different Temperature, Humidity, Wind, and Travel Controls

	Temperature Cntrls:	Bins	Bins	$Avg.^2$	$Avg.^2$	Min^2 & Max^2	Min^2 & Max^2
	Humidity Cntrls:	Bins	Bins	Heat Risk	Heat Risk	Heat Risk	Heat Risk
	Wind & Travel Dist?:	Y	N	Y	N	Y	N
Mean Ozone (10 ppb)	Men	**0.0107****	**0.00999****	**0.0167*****	**0.0173*****	**0.0134*****	**0.0127*****
		(0.00479)	(0.00479)	(0.00438)	(0.00440)	(0.00448)	(0.00449)
	Women	**0.00946****	**0.00899****	**0.0155*****	**0.0159*****	**0.0138*****	**0.0134*****
		(0.00447)	(0.00448)	(0.00418)	(0.00418)	(0.00427)	(0.00426)

Notes: Robust standard errors in parenthesis are two-way clustered at the meet and athlete levels. The results in column 1 match the main results presented in Table 2.6. All regressions include controls for home field advantage, experience (year-of-eligibility fixed effects), practice and meet importance (week-of-season fixed effects), venue characteristics (venue fixed effects), gender and event-specific peculiarities of the normalized outcome measure (event fixed effects), and unobserved athlete heterogeneity (athlete fixed effects). Measures of ambient pollution are averages of ambient levels across meet days.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.10: Estimates by Endurance Event

	800m Run		1,500m Run		3,000m Steeplechase		3,000m Run		5,000m Run		10,000m Run	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Meet Ozone Level	**0.00872**	**0.00887**	**0.0514**	**-0.0406**	**0.591**	**1.078***	**-0.0971**	**0.168**	**1.765*****	**3.054*****	**2.525**	**10.98****
(10 ppb O_3^{Meet})	(0.05)	(0.07)	(0.11)	(0.16)	(0.36)	(0.60)	(0.90)	(1.02)	(0.56)	(0.81)	(2.51)	(4.71)
Mean Event Outcome	121.94	149.10	257.76	307.36	603.84	720.94	557.50	669.16	961.85	1158.47	1990.88	2390.04
SD Event Outcomes	7.99	13.37	21.67	26.89	43.78	61.83	35.26	51.96	64.36	95.91	137.63	211.64
Observations	28,081	28,620	29,246	27,422	8,674	7,348	1,957	2,944	16,532	14,263	4,335	3,269
Adjusted r^2	0.867	0.906	0.92	0.916	0.904	0.902	0.875	0.879	0.891	0.903	0.906	0.927
Number of Athletes	6822	7103	7975	7145	2193	1890	775	1118	5009	4255	1522	1158

Notes: Robust standard errors in parenthesis are two-way clustered at the meet and athlete levels.All regressions include controls for wind assist, travel distance, temperature ($5^{\diamond}F$ bin indicators), dew point ($2.5^{\diamond}F$ bin indicators), home field advantage, experience (year-of-eligibility fixed effects), practice and meet importance (week-of-season fixed effects), venue characteristics (venue fixed effects), gender and event-specific peculiarities of the normalized outcome measure (event fixed effects), and unobserved athlete heterogeneity (athlete fixed effects). Measures of ozone are ambient levels across meet days.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.11: Meet and Training Ozone Levels Interacted

	Endurance Events	
	Male	Female
Meet ozone Level	**0.0434*** **	**0.0406*** **
(10 ppb O_3^{Meet})	(0.01050)	(0.01070)
Meet & Home ozone Levels Interacted	**-0.0117*** **	**-0.0112*** **
($O_3^{Meet} \times O_3^{Home}$, 10 ppb both)	(0.00318)	(0.00354)
Observations	104,564	98,298
Athlete #	18,618	17,311
Meet #	1,684	1,678
Adjusted r^2	0.861	0.888

Notes: Robust standard errors in parenthesis are two-way clustered at the meet and athlete levels.All regressions include controls for wind assist, travel distance, temperature ($5^\diamond F$ bin indicators), dew point ($2.5^\diamond F$ bin indicators), home field advantage, experience (year-of-eligibility fixed effects), practice and meet importance (week-of-season fixed effects), venue characteristics (venue fixed effects), gender and event-specific peculiarities of the normalized outcome measure (event fixed effects), and unobserved athlete heterogeneity (athlete fixed effects). Measures of O_3^{Meet} are ambient levels across meet days, while measures of O_3^{Home} are annual averages from the nearest monitor. O_3^{Home} does not enter the specification directly because of athlete and year of eligibility fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.12: Effect of O_3^{Meet} by O_3^{Home} Level

Bins by O_3^{Home} Level	Endurance	
	Male	Female
Below 20 ppb	**0.0234***** (0.01)	**0.0125*** (0.01)
20-22.5 ppb	**0.0209***** (0.01)	**0.0171***** (0.01)
22.5-25 ppb	**0.0139**** (0.01)	**0.0152***** (0.01)
25-27.5 ppb	**0.00856*** (0.01)	**0.0111**** (0.00)
27.5-30 ppb	**0.0108**** (0.01)	**0.00647** (0.00)
30-32.5 ppb	**0.00743** (0.01)	**0.00516** (0.01)
32.5-35 ppb	**0.00842** (0.01)	**0.00807** (0.01)
35-37.5 ppb	**-0.00518** (0.01)	**-0.00692** (0.01)
37.5-40 ppb	**0.00495** (0.01)	**-0.0043** (0.01)
≥40 ppb	**0.00619** (0.01)	**-0.00508** (0.01)
Observations	104,564	98,298
Athlete #	18,618	17,311
Meet #	1,684	1,678
Adjusted r^2	0.861	0.888

Notes: Robust standard errors in parenthesis are two-way clustered at the meet and athlete levels. All regressions include controls for wind assist, travel distance, temperature (5$^\diamond F$ bin indicators), dew point (2.5$^\diamond F$ bin indicators), home field advantage, experience (year-of-eligibility fixed effects), practice and meet importance (week-of-season fixed effects), venue characteristics (venue fixed effects), gender and event-specific peculiarities of the normalized outcome measure (event fixed effects), and unobserved athlete heterogeneity (athlete fixed effects). Measures of O_3^{Meet} are ambient levels across meet days, while measures of O_3^{Home} are annual averages from the nearest monitor.

*** p<0.01, ** p<0.05, * p<0.1

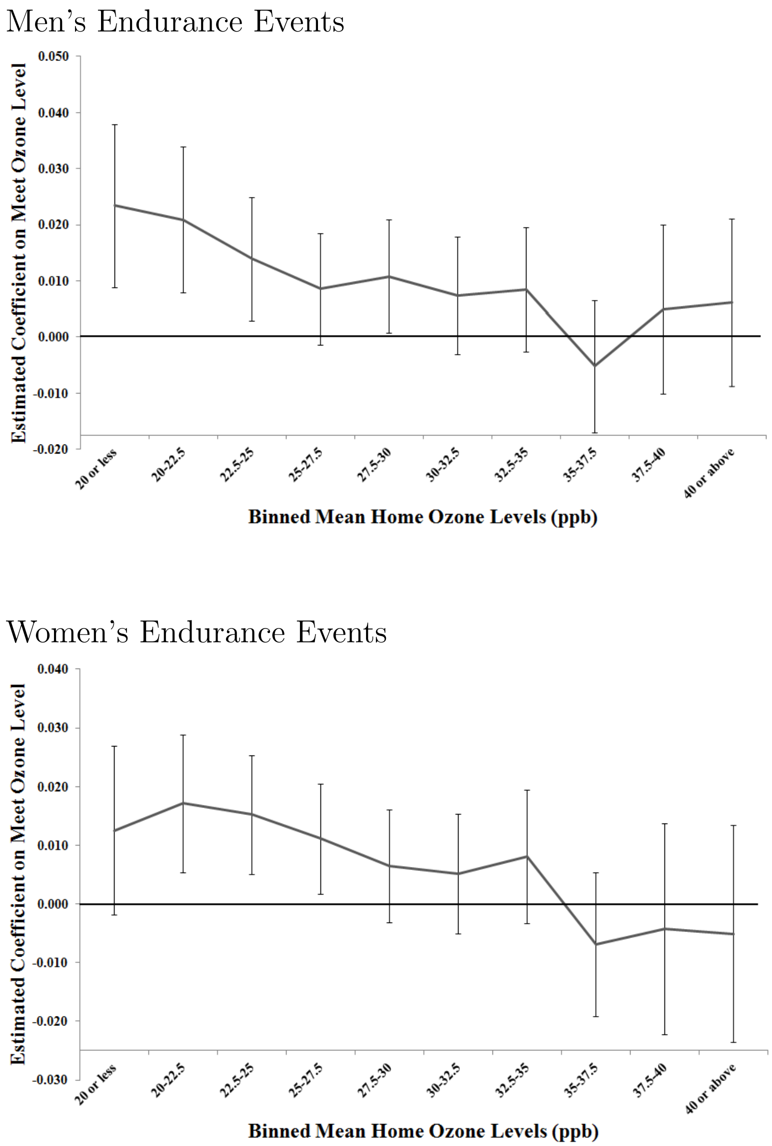


Figure 2.3: Effects of O_3^{Meet} by Home Ozone Level - Endurance Events

Notes: Charts include all outdoor endurance events linked to weather and pollutant monitors for control factors.

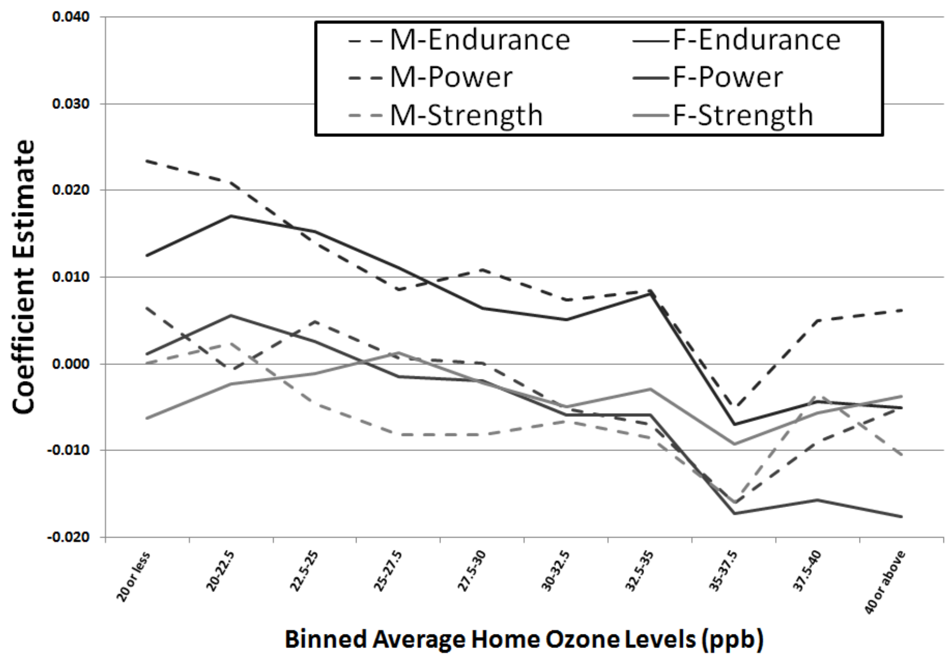


Figure 2.4: Effects of O_3^{Meet} by Home Ozone Level

Notes: Chart includes all outdoor events which could be linked to weather and pollutant monitors for all factors of interest.

2.10 Appendix Table

Table 2.13: Measures by Pollutant Under the EPA's NAAQSs

	NAAQS Unit	Regulated Daily Measure (Used as levels in this paper)	NAAQS Level	Regulated Form of Daily Measure
CO	ppm	max of 8-hr moving avg.	9 ppm	$\leq$ 1 per year
SO_2	ppb	max of 1 hr avg.	75 ppb	99th percentile over 3-years must be $\leq$level
NO_2	ppb	max of 1 hr moving avg.	100 ppb	98th percentile over 3-years must be $\leq$ level
O_3	ppm	max of 8-hr moving avg.	0.075 ppm	3-year avg. of 4th highest annual daily max
$PM_{2.5}$	$\mu g/m^3$	24-hour avg.	35 $\mu g/m^3$	98th percentile over 3-years must be $\leq$level

Notes: Information taken from EPA website: http://www.epa.gov/air/criteria.html. The form of the regulated daily measure is not used in the current investigation.

Chapter 3

Effects of Short Term Measures to Curb Air Pollution: Evidence from Santiago, Chile

3.1 Introduction

Over the past 50 years, an increasing number of national, regional, and city governments around the world have taken deliberate actions to address local air pollution. Although there has been recent discussion about global transport of pollutants, the share of local ambient air pollution attributable to foreign sources is generally quite small (see, for example, EPA 2010), and local air quality is still largely determined by economic development, topography and weather patterns in the immediate area. Unlike global pollution issues, such as the deterioration of the Ozone layer or global climate change, unilateral local action has the potential to address many local air pollution problems. This has led to a variety of government responses which vary greatly in their methods and levels of success. Broadly, such policies can be thought of as either measures that tackle pollution in the long run (switching to cleaner sources of energy, mandatory car emissions standards, etc.) or measures that tackle pollution problems in the short run (driving restrictions in *response* to current air quality, temporarily shutting down or reducing usage of stationary emissions sources, etc.). While the literature examining the impacts of long run pollution abatement policies is rich (see for example: Davis, 2008; Auffhammer and Kellogg, 2011; Chay and Greenstone, 2003), few studies empirically examine the effects of short term measures to address poor air quality (notable exceptions are Cutter and Neidell, 2009; Neidell, 2009; Graff Zivin and Neidell, 2009).

Quantifying such effects is important. It has been shown that even short term exposure to increased levels of air pollution can impose significant costs on human health (Graff Zivin and Neidell, 2012; Neidell, 2009; Currie et al., 2009; Schlenker and Walker, 2011), and many large cities in the developed and emerging world continue to struggle with intense bouts of high pollution. For instance, both Paris and Delhi have recently been in the news for poor air quality (BBC News, March 2014; New York Times, January 2014). More notably Beijing has suffered numerous (highly visible) air pollution spikes, which have prompted political leaders to address the issue. However, even in that instance, most of the debate has surrounded long term changes like relocating power plants and updating cars and buses to pollute less (China Daily, March 2013). While long term solutions are certainly important for the overall reduction of pollution levels, this paper shows that even short run responses to peak pollution events can have positive impacts in the near term and can perhaps mitigate some of the societal costs of elevated pollution levels.[1] In some cases, measures like those explored in this paper could even serve as a stop gap until more permanent air quality management strategies are developed and implemented.

Santiago, Chile is particularly susceptible to poor air quality.[2] Prior to the start of significant government interventions, PM_{10} levels over $300\mu g/m^3$ were not uncommon and occasionally levels of more than $500\mu g/m^3$ were measured within the city. Considering that the World Health Organization guideline for PM_{10} is a 24-hr mean value of $50\mu g/m^3$, it is unsurprising that Santiago was widely known for its poor air quality (WHO 2011). In response to health concerns, and growing public discontent with the air quality in the city of Santiago, the Government of Chile instituted a string of policies in the late 1980s and early 1990s to address worsening air pollution. These actions culminated in the 1997 publication of the Plan de Prevención y Descontaminación Atmosférica (in English: Plan to Prevent and Reduce Air Pollution, and hereafter referred to as the "PPDA"), which laid out a governmental approach

[1]Examples of costs imposed by air pollution include: degradation of asset valuations (Chay and Greenstone, 2005), diminished quality of life (Luechinger, 2009), reduced experiential values (Carson et al., 1992), property and ecosystem damage (Likens et al., 1996), and reductions in economic output (Ostro, 1983; Graff Zivin and Neidell, 2012).

[2]Throughout this paper we use the term "poor air quality" interchangeably with high levels of air pollution and/or PM_{10}. PM_{10} is one of the main contributors to poor overall air quality in Santiago and levels of PM_{10} solely dictate the short term responses we examine. PM_{10} is a measure of particulate matter in the ambient air with diameter of 10 micrometers or less. PM_{10} levels are usually reported in units of micrograms of such particulates per cubic meter of ambient air, or $\mu g/m^3$.

explicitly intended to reduce air pollution in the Santiago Metropolitan Region and mitigate the negative health effects of air pollution exposure among the population. While the PPDA included a number of provisions intended to address air pollution in the short and long terms, central among them was a program under which the government would publicly identify days on which levels of air pollution were expected to exceed certain threshold levels, and flag such days according to a tiered labeling system. Announcement of the higher levels of such "Environmental Episodes" (hereafter, simply "Episodes") were accompanied by mandatory restrictions on driving, the shutdown of certain major stationary emitters (see table 3.6 for specific shutdown requirements), and a number of other protocols intended to reduce ambient pollution levels and shield the public from exposure (detailed in table 3.5). The analysis in this paper focuses on the effectiveness of the entire suite of policies implemented under Episode announcements in reducing air pollution and improving health outcomes amongst the population.[3]

One of the main challenges in an empirical examination of such policies is isolating the impact of the policy from other factors that might also be driving pollution levels in the short run. Since Episodes are announced when pollution levels are well above average levels, it is hard to determine whether a subsequent observed drop in air pollution levels is in response to the enacted interventions or simply the recession from a natural maximum which would have occurred regardless of policy actions (i.e.- mean reversion). While it is difficult to pin down the causal impacts of such policies in the long run, we can do better in the short run by using the fact that Episodes are only announced on days when pollution levels are forecast to reach a certain level. By comparing outcomes in the days following an Episode to outcomes on days following a similar high pollution event when no Episode was announced, we are able

[3]We will use the term Episode to include both the informational and implementational aspects of the policy. Thus, for our purposes, an "Episode announcement" encompasses both the identification of a high air pollution day, and the automatic counteractive measures taken under such circumstances.

to ascertain the effectiveness of Episodes in reducing pollution levels and air quality related mortality. Moreover, we exploit the fact that the Episode announcements were not fully implemented before the enactment of the PPDA in 1997, giving our design a period of time (prior to the policies of interest) when peaks in air pollution were effectively left unaddressed. Our empirical methodology consists of propensity score matching (to identify days in the pre-PPDA period with similar pre-treatment covariates to Episode days in the post-PPDA period) followed by a difference-in-differences analysis (outcome variables on days before and after the Episode announcement are differenced across the pre and post-PPDA regimes) to get at the causal effects of Episodes on pollution and mortality levels. Finally, we are aware that when examining health impacts of such policies, avoidance behavior becomes an important empirical confounder (Neidell, 2009; Moretti and Neidell, 2011). Our empirical strategy, while effective at dealing with mean reversion in pollution levels, does not separately identify the health impacts of avoidance behavior. Hence, we interpret our estimated mortality effects as the full impact of the policy on mortality, capturing deaths avoided through both lower levels of air pollution and avoidance behaviors undertaken by a more informed public.

We find that the metropolitan area of Santiago has been able to effectively address high-PM_{10} levels on a short term basis through the use of Episode announcements. Days following an Episode announcement experience significantly lower levels of air pollution compared to similar days with no announcement. This is most starkly illustrated in panel A of figure 3.1, where we see that day 2 after an Episode announcement has approximately 25% percent lower PM_{10} levels than we would expect had the Episode not been implemented. Hence, in the short term, Episode announcements as executed after the PPDA, appear to have had a significant impact in improving air quality in Santiago. Our results on mortality are generally supportive (although not always statistically significant) of the idea that Episodes reduce deaths and in particular deaths among the elderly and deaths attributed to respiratory ailments. We find that 3 days after an Episode announcement, there are approximately 15 fewer (cumulative) deaths above the age of 64 (or 3.8 fewer deaths per 100,000 over-64 population), and more than a quarter of this reduction is due to decreases in deaths due to respiratory causes. This is in line with the idea that poor air quality harms health via respiratory illness. Such results also corroborate other findings demonstrating that even short term air pollution exposure can significantly impact human health and well-being (Graff Zivin and Neidell, 2012; Neidell, 2009; Currie et al., 2009; Schlenker and Walker, 2011).

This study adds an important dimension to the broader literature examining the intersection of air quality and health. While studies have shown that air quality matters for health and other outcomes of interest to economists - like productivity and school attendance (Graff Zivin and Neidell, 2012; Currie et al., 2009) - these papers do not directly tackle the impacts of air quality policies. Most of the studies examining the impact of air quality policies on health have examined long term policies like the role of the Clean Air Act (Chay and Greenstone, 2003; Sanders and Stoecker, 2011) or the NO_x Budget Trading Program (Deschenes et al., 2012). Hence, this paper's main contribution is a novel analysis of the environmental and health impacts of short term policy measures to tackle air quality.

Our paper also adds to the literature measuring the effects of localized governmental policies aimed at curbing air pollution. Most of these studies and policies focus on driving restrictions and have found mixed results. Using data from São Paulo, Bogotá, Beijing, and Tianjin, Lin et al. (2013) find that driving restrictions may reduce the incidence of extremely high concentrations of air pollutants, but that such restrictions do not, on average, significantly improve overall urban air quality. Davis (2008) concludes that one-weekday-per-week driving restriction in Mexico City did not significantly improve air quality in that city, and finds that the restriction likely drove the purchase of additional (often older and dirtier) vehicles by households to get around the restrictions. In addition to examining the environmental impacts of similar but short term policies in Santiago, we extend our analysis to examine the health impacts of such policies. Quantifying this natural externality of pollution abatement policies, especially in the context of a developing country, is a novel addition to the literature.[4]

Finally, our paper's findings are in line with a recent paper also examining the short term effectiveness of Episode announcements in Santiago. Troncoso et al. (2012) find that the announcement of Episodes leads to significant reductions of particulate matter, CO, NO_x, and O_3, while having no effect on SO_2. Our paper adds to the Troncoso et al. (2012) analysis in a number of important ways. First, a broader time period - including the initial introduction of the PPDA - is examined, allowing us to take advantage of greater overall variation in pollution. Second, we explicitly account for mean reversion of ambient pollutant levels and the fact that Episode announcements are likely to correspond with natural peaks in pollution levels. Third, we address the multi-day effects of Episode announcement, which are not discussed by Troncoso et al. (2012), but prove to be a significant portion of the overall benefit of such announcements. Finally, we examine and are able to identify mortality impacts (albeit with low power in some cases) of the Episodes program, something that Troncoso et al. (2012) does not consider, but is clearly of great interest.

[4]Related to short term measures and pollution alerts, a few papers have examined the impacts of the provision of information regarding air quality via public announcements. For example, Cutter and Neidell (2009) find changes in public transportation use and traffic volume in response to "Spare the Air" campaigns (a voluntary campaign put in place on high ozone days in Northern California). Neidell (2009) finds that smog alerts affect attendance in outdoor facilities, and shows that a correct accounting for this behavioral response increases the estimated impact of ozone on asthma related hospitalizations. Using a similar source of information (successive days of smog alerts in Southern California), Graff Zivin and Neidell (2009) show that any behavioral response to alerts on the first day nearly disappears by the second sequential day of alerts. Note that most of these informational alerts are not accompanied by restrictions on polluting sources or other measures to address pollution levels directly. The fact that behavioral responses to information are important is a point we return to later in the text.

3.2 Background

Santiago's geographic position in a basin at the foot of the Andes mountain range leads to the frequent occurrence of temperature inversion layers over the city. Such inversion layers reduce vertical atmospheric mixing, thereby trapping pollutant emissions near the ground (Prendez, Alvarado and Serey, 2011). Thermal inversions are common in summer as well as in winter; however, in the winter months (April-August), the inversion layers tend to be much closer to the ground, consistently leading to increased levels of air pollution at ground level (Gramsch et al., 2006; Rutllant and Garreaud, 1995). Since the 1960s Santiago has regularly suffered from periods of air pollution far in excess of levels considered healthy for the region's ever growing population (now estimated at 7 million people in the Metropolitan Region).

Beginning in the late 1980s, the government of Chile implemented a string of policies intended to address air pollution in the greater Santiago area. This included the establishment of an automated network of pollution monitors in 1988, the mandated annual inspection and ranking of stationary emissions sources based on the concentration of pollutants emitted in 1990, and the requirement that all new cars have catalytic converters in 1992 (Chilean Ministry of the Environment, 2007). Figure 3.2 shows that PM_{10} levels have fallen significantly and consistently as a result of these, and other programs.

The policies that are the focus of this paper were created in 1990, and involve the Chilean Government preemptively identifying days during the winter months (March-September) on which air pollution is anticipated to be particularly severe. Such days, known as Episodes, are dubbed "Pre-Emergency" Episodes if forecast PM_{10} concentrations exceed $240\mu g/m^3$, and "Emergency" Episodes if PM_{10} levels over $330\mu g/m^3$ are anticipated (Supreme Decree 32, 1990). For the purposes of our analysis, it is important to note that although the policy of identifying and announcing Episodes was technically established in the early 1990s, evidence suggests that it was not vigorously implemented until much later. Figure 3.3 shows the shift from lax execution of Episodes coincided roughly with the 1997 implementation of the PPDA.[5] In the pre-PPDA period, we see 148 days when PM_{10} was in excess of $240 \mu g/m^3$ (i.e.- an Episode was warranted under the law), but less than 40% of these days were announced as Episodes. Conversely, after the passage of the PPDA, the government announced Episodes on nearly all high-pollution days.[6] We conclude that the Episodes policy practically came into force beginning in 1997 with the PPDA, and proceed with our analysis, treating 1997 as the working implementation date of the policies of interest.

This approach assumes that even the relatively small number of pre-PPDA Episodes that *were* announced, were ineffective and therefore days on which pre-PPDA Episodes occurred were no different from other pre-PPDA days. Figure 3.5 supports this assumption visually, and when we assess the potency of the pre-PPDA Episodes using our empirical framework, we find that such Episodes were relatively ineffectual (results not presented). We thus proceed by treating all pre-PPDA days as if they were "untreated" by the Episode program of interest. For further discussion of pre-PPDA Episodes and a robustness check in which they are excluded from the control group see Section 3.4.1. Finally, it is worth noting that if the pre-PPDA Episodes were in fact effective, our strategy would result in *conservative* estimates of the effectiveness of short term environmental regulations post-PPDA.

Episode implementation proceeds as follows. Each day, a group of

[5]The PPDA also introduced a new, lower level of Episode implemented when PM_{10} concentrations were expected to exceed $195 \mu g/m^3$. The new level is called an "Alert" Episode, and is generally omitted from the analysis of this paper because it was often announced for many days in a row (making identification of the effects of any one announcement impossible) and came with only minimal additional driving restrictions and no implications for stationary emitters. For the purposes of this paper, the term Episode will only refer to Pre-Emergency and Emergency announcements.

[6]To drive this point home, note the similarities between 1996 v. 1998 and 1994 v. 1999 in figure 3.3. Each pair had comparable numbers of high-PM_{10} days, yet we see in figure 3.3 that the years after the implementation of the PPDA (namely: 1998 and 1999) have dramatically more Pre-Emergency and Emergency Episode announcements. This suggests that high air pollution days were met with quite different government actions before versus after the implementation of the PPDA.

air quality forecasters examine current weather and pollution conditions, and forecasts for the subsequent days. If these professionals conclude that one of the PM_{10} concentration thresholds described earlier will be breached during the following day, it is recommended that an Episode be announced for that day. A number of officials then sign-off on the decision, and the Episode is officially announced publicly by 8pm the evening prior to its implementation (Saide et al., 2011). As the decision to announce an Episode (or not) is made based on information available prior to the start of the Episode, we assume that such announcements are exogenous to the ultimate effectiveness of the Episode itself. There is no indication that announcement decisions are impacted by any measure that would be expected to predict the effectiveness of Episode enactment.

The restrictions and governmental actions that go into force automatically upon an Episode announcement differ by Episode level (see table 3.5 for details), but the main thrust of both Pre-Emergency and Emergency Episodes include tighter driving restrictions (some are already in force seasonally), the mandated operational cessation of a share of stationary emissions sources, and an informational campaign designed to get the word out to the public about the Episode.[7] With the exception of driving restrictions - which are only in effect from 7:30am to 9pm - Episode protocols go into force at midnight on the day of the Episode (see table 3.6 for further discussion of protocol timing and methods). In addition to the severe driving restrictions and mandated shutdowns of large shares of stationary emissions sources, a number of other protocols come into force upon the announcement of an Episode. These include extra street sweeping, cancellation of physical education in schools, and restrictions on the use of some residential heating fuels. Additionally, police presence is increased to enforce the restrictions applied on Episode days. As public knowledge and air pollution levels are contemporaneously impacted by

[7]Note that only two Emergency Episodes occurred during the post-PPDA period. We therefore focus on the impacts of Episodes generally, rather than trying to determine effects separately by Episode level.

an Episode, we are unable to attribute effects to specific protocols, or separate the effects of avoidance behavior from other ameliorative impacts of Episode announcement. Therefore, empirically we examine the gross impacts of the Episode policy package, including both public announcement and prescribed government actions.

3.3 Empirical Approach and Data

In order to examine the short term effects of Episodes, we must develop an approach that allows us to control for mean reversion. Because Episodes are typically announced at times when air pollution is above mean concentrations, on average we would expect air pollution levels to fall on subsequent days whether or not any (effective) actions were taken. Failing to take account of this fact - as an event study or regression discontinuity approach would - will lead to upward bias in our estimated effects. Additionally, other pre-Episode factors, including weather shifts, may lead to differential changes in pollution levels that are unrelated to Episode announcements. In order to account for mean reversion (and perhaps other effects that might change outcomes, but are unrelated to Episode announcements) we will utilize a difference-in-differences approach, comparing changes in outcomes from before to after an Episode to changes in outcomes from before to after another (similar) day when no Episode was announced. In the remainder of this section, we briefly describe the data used in our analysis and then lay out our empirical methods.

3.3.1 Data

The empirical analysis in this paper relies on a panel data set which covers 1989-2008, and was created by merging day-level data from a number of administrative and observational sources. The data set is built upon the foundation of observational measures of Santiago PM_{10} concentrations, which were collected by the MACAM 1 (pre-1997) and MACAM 2 (post-1997) monitor networks and maintained by Chile's Ministry of the Environment. When the original six monitors (there are now nine) in the MACAM 2 network replaced the five monitors in the MACAM 1 network, they were deliberately spread more widely throughout the city (see figure 3.4 for a comparison of monitor locations under the MACAM 1 and MACAM 2 networks), with placements intended to capture traditional hotspots and provide observations on representative pollution levels (Gramsch et al., 2006). Due to the adjustment of monitor locations between the networks, only three sites were monitored over the entire period of our study. These consistently monitored locations are: Parque O'Higgins, La Paz (referred to as "Independencia" in some sources), and Las Condes. In order to ensure the most consistent comparison of PM_{10} levels before and after 1997, our primary analysis uses only monitors at these three sits. Monitors at all three locations were consistently functional during the period of study and all the pollution monitor data used in this paper were collected and maintained by the Ministry of the Environment. PM_{10} data are aggregated to the average daily level by station, and the "citywide" mean across stations is the focus of our analysis.[8] See figure 3.2 for a plot of PM_{10} levels in Santiago over the period of our study, and section 3.4.1 for a robustness check of our monitor selection.

As weather conditions are expected to covary with many of the outcomes of interest in this study, observational weather controls are of critical importance. Daily minimum, maximum, and mean levels of a large number of meteorologic variables (temperature and precipitation for example) are calculated from hourly observations reported in the Summary of the Day data series from the U.S. National Climatic Data Center (NCDC). Daily means of weather variables are then merged with the daily pollution data for the entire period from 1989-2008.[9] Administrative data on the dates and levels of

[8] Due to the centrality of PM_{10} levels in our examination, days for which PM_{10} data are not available from any of these three stations are omitted from our analysis. This criterion leads us to omit 185 days in the pre-PPDA period and 17 days in the post-PPDA period (all in 1997) from the matching analysis.

[9] Observational weather data are taken from the Pudahuel station (#855740). Weather

Episode announcements from the Santiago Metropolitan Region's Ministry of Health are also added to the panel.

Unfortunately, mortality data for the Santiago Metropolitan region are only available starting in 1992, and cause-of-death information is not available until 1994. Mortality data were obtained from the Chilean Ministry of Health's Department of Statistics and Health Information, and include information on each death in Chile for the period 1994-2008, including date of death, age of the deceased, and International Classification of Diseases (ICD) codes for primary and secondary causes of death.[10] Only data on age and date of death are available for the years of 1992 and 1993. The mortality data are aggregated to the daily level for the Metropolitan Region and added to the panel. It is important to note, that the limited temporal scope of the available mortality data means that we will have slightly fewer matched observations for analyses involving mortality, and fewer still for those that rely on cause-of-death information.

Table 3.7 presents mean values for a number of variables used in the analysis.

3.3.2 Identifying Treatment (Episode) Days

In order to examine magnitudes of an Episode's impact on air quality and mortality outcomes, it is necessary to observe an Episode "treatment" in isolation from confounding factors and additional treatments. A factor that could prove problematic for our analysis is the clustering of Episode announcements. Often, Episodes would be announced repeatedly over a sequence of days when air pollution levels were particularly high. In order to avoid confounding the impacts of one Episode with those of another, our analysis focuses

monitor data from this station are missing for only three days during the study period: 9/26/1992, 1/10/1999, and 9/23/2003, none of which are near enough to an Episode or high PM_{10} event to affect the analysis.

[10] ICD-9 codes are used in the data before 1997, and ICD-10 coding is used thereafter. We account for the changing classification.

on Episode days which are neither preceded nor followed by another Episode for a five day period. Thus, under our separation criteria, if an Episode is announced on Day 0, no other Episodes may occur between Day -5 and Day 5.

While this approach limits the number of Episodes used in the analysis to 35 (a total of 91 post-PPDA Episodes were announced during the period examined), it helps ensure that the estimated impacts are appropriately attributed. Such selection criteria, however, might select Episode events that are particularly effective. At the very least, the results of Graff Zivin and Neidell (2009) suggest that the power of Episodes to induce behavioral change via information provision may quickly wane with the announcement of additional Episodes in a relatively short time period. Nevertheless, the spirit of our results is generally maintained under other separation criteria and with the inclusion of *all* Episodes in the analysis (See section 3.4.1 for further discussion and robustness checks).

3.3.3 Identifying Appropriate Counterfactuals

Given that the impact of mean reversion on pollution levels is likely to be non-trivial, simple event study or regression discontinuity approaches, comparing outcome variables before and after an Episode, will not be sufficient. Instead, a comparison group is needed. Since the focus of this investigation is on the short term impacts of Episodes as they were structured under the PPDA protocols (i.e. 1997 and after), it may seem that the obvious comparison group is the days on which Episodes were announced before the PPDA. We are, however, interested in assessing the impacts of Episode announcements, and not the differential impacts of Episodes following the PPDA. Thus, limiting our comparison group to pre-PPDA Episodes does not serve the needs of this analysis. Also, we showed earlier that implementation of Episodes was both inconsistent and ineffective prior to 1997, which suggests that pre-PPDA

Episodes, as a group, may not be the best comparisons for the evaluation of post-PPDA Episodes. Conversely, figure 3.3 shows that the Chilean government's zealous implementation of Episodes in 1997 and after likely robs us of closely comparable non-Episode days in the post-PPDA period.

We thus look in the pre-PPDA period (when many days likely "should" have had Episode announcements) for days that are comparable to Episode days in and after 1997. As most days in the period from 1989-1996 are not similar to days after 1996 on which Episodes were announced, a direct comparison of mean outcomes would lead to high levels of upward bias in the magnitude of our results (Abadie, 2005). Instead, we use matching techniques to identify days from the pre-PPDA period that are "similar" to Episode days in the period after the PPDA. "Similarity" in this case will be based on pre-treatment characteristics of the days leading up to an Episode. Unfortunately, even using only straightforward covariates such as PM_{10} and weather variables for the pre-Episode days, it is not possible to find exact matches amongst days in the pre-PPDA period. Following Rosenbaum and Rubin (1983), we use propensity scores to reduce the dimensionality of the matching problem.

Matching on propensity scores provides a quantitative approach to linking post-PPDA-Episode days to similar days on which an Episode of interest was not announced. Each Episode in the post-PPDA period is matched, based on a Logit-generated propensity score, to a number of days in the pre-PPDA period. The propensity score can be thought of as an estimated probability that a given day would have had an Episode announcement. The value of the score is generated by estimating a Logit model on predetermined characteristics of days in the post-PPDA period on which Episodes were announced, and using the estimated coefficients to predict the probability each day in the period of examination would have had an Episode announcement. The following specification of the Logit model is estimated in order to generate propensity scores for days in the sample:

$$y_t = \alpha + \sum_{j=1}^{5} \left(\beta_j * PM10_{t-j} + \mathbf{X'_{t-j}} \gamma_j \right) + \mathbf{DOW'_t} \delta + \mathbf{month'_t} \theta + \varepsilon_t \tag{3.1}$$

In this specification, the left-hand side variable, y_t, is simply an indicator variable that takes on the value of one if an Episode was announced on day t in the post-PPDA period. $PM10_t$ captures the mean PM_{10} concentration on day t, while X_{t-j} is a vector of observed weather variables j days before day t including mean temperature, average windspeed, and precipitation. $\mathbf{DOW_t}$ and $\mathbf{month_t}$ represent day-of-week and month-level fixed effects (note that γ_j, δ, and θ are vectors of coefficients), and the error term is written as ε_t. Pre-Episode conditions are captured in the lagged values of the pollution and weather variables incremented by j. We match based on the conditions in the 5 days preceding an Episode.

Weather variables for the day of the Episode or potential match day are excluded because their levels are not determined at the time of the Episode announcements.[11] Day of the week is included as it likely captures some emissions information that may improve match quality, and although it is not technically a factor that was to be considered by authorities, the day of the week may impact the government decision of whether to announce an Episode or not. Month dummies are included to capture both seasonal variation in weather patterns, which likely contribute to fluctuations in air pollution, and seasonal variation in Episode announcements connected to attitudes within government.

Once the above Logit model is estimated, each post-PPDA Episode and pre-PPDA day is plugged into the model, and the resulting predicted value of y_t is the propensity score for day t. In our headline result each post-PPDA

[11]While the inclusion of day-of weather variables might be justified because they are unaffected by treatment, or because they serve as proxies for forecast values that officials may have consulted, such inclusion does not markedly change our results.

Episode is then matched to the five pre-PPDA days with the most similar propensity scores. This method is known generally as the five Nearest Neighbor approach.[12] Matching each Episode to multiple pre-PPDA days reduces the variance of our estimates, while limiting the number of matches reduces the possibility of using poor matches in our analysis. We enforce common support on propensity scores between the Episode and matched groups, which leads us to drop one Episode day from our analysis because its propensity score is above that of any comparison day. Enforcing common support is important theoretically and empirically to ensure that matches are in fact similar (Heckman et al., 1999).

Table 3.1 presents mean values of the matching variables for the control and treatment groups.

3.3.4 Difference-in-Differences

Now that we have identified a set of matched days in the pre-PPDA period, we have an appropriate "control" group against which to compare the outcomes of our "treatment" group of days with a post-PPDA Episode announcement. Since the goal of this exercise is to identify the effects of an Episode on several different outcomes, we will compare differences over time in the outcome variables across the date of the Episode (or Episode matches) between the treatment and control groups. Using a difference-in-differences (hereafter: "DID") strategy helps ensure comparability of outcomes on similar days from different periods given the different air quality and mortality characteristics between the periods.[13] The DID approach controls for long-

[12]See section 3.4.1 for a demonstration of the robustness of our results to the use of alternative matching procedures including Caliper and Kernel based methods. Additionally, note that we match with replacement in order to maximize the quality of our matches, and thus reduce the bias of our estimates, though possibly at the cost of higher variance (Smith and Todd, 2005; Abadie and Imbens, 2005). This means that some pre-PPDA days are matched to multiple Episodes. When this occurs, such days are appropriated higher weightings in the regression analyses.

[13]Mean PM_{10} levels are different in different years, thus "spikes" in PM_{10} are different between years, both in absolute levels and in relation to the thresholds warranting Episode

term mean changes in our outcome variables[14] better than a time trend in the propensity score. Additionally, Smith and Todd (2005) find that DID estimators better address population and measurement method mismatch between treatment and control groups than do matching estimates based on comparisons of levels.

Conceptually, our first difference (of the DID) will be changes in pollution levels or number of deaths (our outcome variables of interest) from before to after a pollution event (i.e.- an Episode or matched day). Identification then comes off a comparison of these pre to post-event changes in the outcome variable among the treatment group to the same changes for events in the control group (this is the second difference of the DID). Our first results are the traditional comparison of the means (of these differences) between the treated (Episode) days and matched control days. These results are presented in table 3.2. No additional controls are included at this stage of the analysis because, in theory, propensity score matching effectively controls for all the inputs of the matching procedure.

3.3.5 Regression Analysis

The strict comparison of means relies on the correct specification of the propensity score estimation and perfect matching (i.e.- exact balance in the distributions of all observable and unobservable covariates) between our treatment and control groups. Table 3.1 shows the balance achieved in our observable matching covariates, and demonstrates that our matching procedure has given us good - though not perfect - balance.

In order to control for remaining pre-treatment differences between our treatment and control groups, we implement a "mixed method" like those laid out in Imbens (2004) and Hirano and Imbens (2001), which involves esti-

---

announcements.

[14]Year on year, air quality was improving and the number of deaths in Santiago was growing as the population increased.

mating regressions using the samples and weightings generated in the matching procedure. In addition to controlling for remaining differences between our comparison groups, regression analysis, with proper controls, will address correlations between the matching covariates and our outcomes of interest, thereby potentially increasing the precision and/or reducing the bias of our estimates (Imbens, 2004). Approached from another perspective, we are using propensity score matching prior to a difference-in-difference analysis to better meet the necessary assumptions regarding parallel evolution of paths between the treatment and control groups (Abadie, 2005).

Although we run a number of regression specifications (results from some are presented as robustness checks), our main specification is simply a traditional DID OLS regression assessing the interaction of indicator variables for treatment and being a member of the treated group. The key to this analysis is that it is run, not on the whole 1989-2008 sample, but on observations (i.e.- days) associated with pollution events in the treatment and control groups obtained via propensity score matching. Such an analysis is analogous to regressions estimated following the implementation of a randomized control trial, as additional controls are still useful even when it is believed that treatment and control groups are closely comparable.[15] Below is the specification of our mainline regression analysis which is estimated on six distinct samples (each of which examines Episode effects over a different time-frame, ranging from day-of to five days after announcement) where the coefficient of interest is on the interaction term:

$$\begin{aligned} Y_{it} &= \alpha + \eta * E_i + \delta * P_t + \beta * (P_t * E_i) \\ &\quad + \sum_{j=1}^{5} \mathbf{X}'_{\mathbf{i-j}} \gamma_j + \mathbf{DOW}'_{\mathbf{i}} \delta + \mathbf{month}'_{\mathbf{i}} \theta + \varepsilon_{it} \end{aligned} \tag{3.2}$$

[15]In the context of an RCT, the close comparability of treatment and control groups is achieved through random assignment. While we don't have random assignment here, the goal of our matching procedure is to mimic random assignment of Episode days amongst days warranting an Episode.

This specification is a comparison across event studies, where i is date of the event and t denotes distance of the sample observation (in days) from its associated pollution event.

If an observation's associated pollution event, i, is a post-PPDA Episode, the day is in the treatment group and $E_i = 1$, otherwise (for days before and after events in the control group) $E_i = 0$. Sample days that precede a pollution event will have $t < 0$, and are indicated by $P_t = 0$, while sample days following the event have $P_t = 1$. Each sample includes a before ($t < 0$) and an after ($t \geq 0$) day for every event, and our main results are based on analysis of six distinct samples that include the day just prior to each event ($t = -1$) and the day 0, 1, 2, 3, 4, *or* 5 day(s) after the event.

The outcome variable, Y_{it}, will take on either pollution or mortality measures for day t relative to event i. Similar to the Logit specification, X_{i-j} is a vector of observed weather variables for the jth day prior to event i (with $j \in [1, 5]$), and includes mean temperature, average windspeed, and precipitation. Also, $\mathbf{DOW_i}$ and $\mathbf{month_i}$ remain as day-of-week and month-level fixed effects for the day of the event (with γ_j, δ, and θ again representing coefficient vectors). Our error term is ε_{it}.

3.4 Results

Panels A and B of figure 3.1 graphically demonstrate the essence of our results, showing the average movement of air pollution and mortality through time across the threshold of an Episode announcement in the treatment (post-PPDA Episodes) and control (matched pre-PPDA days) groups. In each of the graphs, the vertical line represents the timing of treatment (i.e.- Episode announcement or matched control day), while the horizontal axis groups days by the amount of time before or after the pollution event.

Visually it is clear that relative to similar days in the pre-PPDA period, implementing an Episode at the peak of pollution levels in the post-PPDA period drastically reduces PM_{10} levels. Panel A also shows that our matching methodology for the five days before an Episode announcement works quite well. Since the matching was done purely based on pollution and weather patterns, the match in panel B for deaths is not as close. However, our matching approach stays true to the official methods by which Episode announcement decisions were supposed to be made. Given the difficulty of capturing political or other qualitative factors that might have gone into Episode announcement decisions in practice, we have opted to keep things as simple as possible, basing our matching only on pollution levels and meteorological conditions prior to Episode announcement.

The simplest quantitative analog of figure 3.1 is the comparison of mean changes across treated and control pollution events. The results from this approach are shown in table 3.2, but are less precisely estimated than we might like as this direct differences-in-means approach assumes perfect balance and attempts to mimic a randomized control trial. In table 3.3 we show that the use of regression controls can account for some of the remaining differences between our comparison groups. Table 3.3 in general verifies the results presented in table 3.2, but the magnitudes are slightly larger (though not statistically different from the estimates with no covariates). To ensure our results are not being driven by the addition of covariates, we show that sequentially adding controls does little to alter the overall pattern of results (see table 3.17). However, given the advantages of such mixed methods as outlined in Imbens (2004), we present the results in table 3.3 as our preferred estimates.

Each cell in table 3.3 represents a different regression. As panel A - figure 3.1 suggests, the effect of an Episode announcement on pollution levels is large and statistically significant. Particularly, we see that the day of an Episode experienced average PM_{10} levels that were approximately $22.5 \mu g/m^3$ lower than would have been expected without the Episode announcement. This is a very striking effect given that the mean city-wide level of PM_{10} on the two days preceding an Episode (or matched day) was around $132\mu g/m^3$ and the mean PM_{10} level on matched days is $131\mu g/m^3$. This suggests that an Episode announcement, along with all the government actions and restrictions such an announcement brings into force, leads to immediate PM_{10} concentration reductions of approximately 17% from anticipated levels, with additional air quality benefits continuing for several subsequent days. The magnitude of this day-of estimated effect compares to the high-end of the estimated effects presented by Troncoso et al. (2012) for Pre-Emergencies, but given that our analysis includes only "stand alone" Episodes, this is to be expected. Given the ongoing troubles that many cities have had with high levels of urban air pollution, these results are particularly important as they suggest that governments can effectively improve air quality in the immediate term through counteractive measures.

In order to assess the short term effectiveness of Episode announcements on health outcomes, we estimate regressions using cumulative daily death counts among the elderly as the outcome variable (in table 3.8, we consider other age groups). Columns 2 and 4 of table 3.3 show that deaths and death rates among the elderly appear to decrease after an Episode announcement. A day after an Episode announcement there are 8 fewer overall deaths of those over the age of 64. Given an average of around 65 deaths per day amongst this group during the days leading up to a high-pollution event (1992-2008), this represents a 6% decrease in cumulative deaths. While no direct comparison of this effect size is available, a recent review article by Anderson et al. (2012) - referencing Samet et al. (2000) - suggests a 0.5% increase in short term, all-cause mortality is associated with a $10\mu g/m^3$ increase in ambient PM_{10}.[16] The reduction in PM_{10} we observe by the day after an Episode is around $37\mu g/m^3$, and hence using those estimates, we would expect an effect of around 1.8%. While our estimates appear to be much larger, consider that

[16]Omori et al. (2003), another study cited by Anderson et al. (2012), finds similar magnitude effects for TSP exposure in 13 Japanese cities.

we examine an older (and perhaps more sensitive) sub-population and that the Samet et al. (2000) study was done in 20 US cities where the most polluted city (Los Angeles) had average PM_{10} levels of only $46\mu g/m^3$, and others had average PM_{10} levels well below $30\mu g/m^3$. For comparison, the average daily mean level of PM_{10} in Santiago during the winter months of our study period was $92.6\mu g/m^3$, hence, the much higher underlying levels of pollution could be an important factor in the different effect magnitudes, as marginal responses of health may be non-linear in ambient PM_{10} exposure. Looking at the cumulative effects on over 64 mortality three days after an Episode, we see a decrease in deaths of around 15, which corresponds to a lower death rate over the period of approximately 4 per 100,000 individuals over the age of 64.

As table 3.4 suggests, we do not see any significant reductions in deaths or death rates for causes that are likely not linked with short term exposure to high levels of air pollution. At the same time, large and significant effects are also largely absent from causes we would expect to react to changes in PM_{10} levels. For instance, table 3.4 shows that deaths due to cancer are not significantly affected as a result of Episode announcements (although the coefficients appear sizable), but also that respiratory deaths among the elderly are barely impacted at a significant level. We also see a decrease in deaths due to accidents, and although not statistically significant, this might be due to the elderly staying indoors during pollution events, thereby reducing the likelihood of fatal falls etc. Taken as a whole, our death-by-cause analysis is largely inconclusive due to a lack of power arising from the lower numbers of deaths within cause-specific sub-samples, and the shorter time period (beginning only in 1994) over which we have cause-of-death data. Our examination of deaths over the age of 64 in our mainline results is less prone to power issues since a larger number of daily deaths occur for that age group (approximately 2/3 of daily deaths occur in those with ages over 64) and our data for age-at-death cover a longer period (back to 1992) than the cause-of-death data.

Our analysis is explicitly short term in nature, and we interpret our results on elderly mortality as including harvesting effects. Harvesting is the idea that such peaks in pollution advance the date of passing for those near death by only a small amount rather than significantly impacting overall death rates.[17] Our empirical design is specifically designed to highlight short term impacts, and we are therefore unable to test whether the Episodes program reduces overall mortality over the longer term. To do so would require data from a comparable city (in terms of pollution patterns and underlying health characteristics) that, over the same period, did not have an Episodes program. Given data limitations and that the goal of this investigation is to evaluate the effectiveness of the Chilean approach to spikes in ambient pollution, we set the quantification of harvesting effects aside for future research.

It is worth noting however, that the harvesting process itself is probably not constant over the period of our study since overall health in Chile had been improving during the 1990s (e.g.- life expectancy rose from 73.6 years in 1990 to 78.3 years in 2005), hence, any change in the underlying harvesting process will likely lead to an underestimate of the effects we document. It is also important for us to note that our results could reflect an overestimate if behavioral responses are larger than expected with the initial introduction of the Episodes system. We investigate this possibility further by redoing our analysis after dropping the first 2 years of the program (omitting 1997 and 1998 from the estimation) and find that our results, presented in table 3.9, are largely unchanged. Hence, *if* underlying health and resilience to pollution peaks changed before and after the PPDA implementation period, we believe that we would still be recovering a lower bound effect.

[17] The concept of harvesting is part of a broad literature across various disciplines. See Fung et al. (2005); Peng et al. (2006); Zeger et al. (1999); Zanobetti et al. (2002) for formal models of harvesting.

3.4.1 Robustness Checks

Our results are robust to a broad range of specification checks. As mentioned earlier, our results are not driven by changing monitors across the pre and post-PPDA regimes. Table 3.10 demonstrates that including all monitors that are available - regardless of when the monitors came online or where they are located - does not drastically change our results. This is not altogether surprising given the lack of variation in pollution across *comunas* within the city of Santiago.[18]

Since the results hinge on the variables we use to match and identify control groups, we explore whether our conclusions are robust to the addition of more covariates in the matching process. In table 3.11, we use 5 days of lags for each of: mean PM_{10}, daily max PM_{10}, wind speed, precipitation, mean temperature, maximum temperature, minimum temperature, atmospheric pressure, dew-point, and the square and cube of mean temperature as covariates in the matching process. As table 3.11 shows, the addition of these covariates limits the number of close matches we can identify, and, via enforcement of common support, the number of Episodes we can examine. Nevertheless, the point estimates are closely comparable to our main results, especially for the mortality outcomes. Table 3.12 presents our headline results, reevaluated with the inclusion of lagged daily mortality from the pre-Episode period in the propensity score and matching procedures. The effects on mortality for those over the age 64 are muted in this specification as we would expect given that this approach essentially controls for differences in deaths along one of the dimensions of our DID analysis. Nevertheless, the estimates remain consistent with our main results.

In table 3.13, we continue to explore the sensitivity of our results to different matching methods. Table 3.13 shows that the choice of matching

[18]Las Condes is a bit of an exception with lower pollution than most of the other comunas in Santiago, but robustness checks confirm that our results are not changed substantially by the exclusion of Las Condes monitor data (results not shown).

method is not consequential to our overall set of results, and table 3.14 shows that including all Episodes in our analysis (whether or not each fulfills our separation criteria) does not lead to estimates which contradict our main findings. Additionally, we investigated a number of weighting schemes for use in the regression analysis presented in main results. One commonly discussed approach is to weight control observations by a function of their propensity score (our main results weight only by matching frequency for those control days which are matched to multiple Episodes). A particular embodiment of this approach is laid out in Imbens (2004) (attributing similar methods to Robins et al. 1995; Robins and Rotnitzky 1995; Robins and Ritov 1997). This weighting scheme was implemented and the results are reported in table 3.15. We find that these estimates are not inconsistent with our mainline results.

A potential concern is that including pre-PPDA Episodes as part of the control group could bias our results. Though, as mentioned earlier, the inclusion of pre-PPDA Episodes, if anything, biases our results towards finding no effect of post-PPDA Episodes. In table 3.16, we exclude all pre-PPDA Episodes from our control group and find that our results are indeed stronger. Nevertheless, due to the common support requirement, higher quality matches lead to more precision through larger samples, so we would like to have as large a matching pool as possible. As figure 3.5 indicates, it is not clear that the pre-PPDA announcements had any impact on pollution levels in the near term, hence, our preferred specification includes pre-PPDA Episodes as part of the control group to ensure the largest possible sample for matching and analysis.

3.5 Conclusion

In this paper we have analyzed the short term environmental and health consequences of pollution Episode announcements in Santiago, Chile. As Episode announcements included temporary restrictions on mobile and stationary emitters, we show that such restrictions are effective in reducing pollution levels, and that the combined informational and pollution-reducing aspects of the Episode measures reduce mortality in the short run.

Going forward, we hope to understand more deeply the precise mechanisms behind the overall effects we see. For example, one aspect we cannot address in this paper is that each Episode is really a “bundle” of interventions rather than one specific intervention. From a public policy standpoint, it would be valuable to separately identify the effectiveness of each restriction.

The bundled approach also makes it harder to analyze the economic costs of such restrictions. While pollution levels and mortality are reduced by these restrictions, such gains must certainly come at a significant cost. In future work, we hope to gain a better understanding of the costs of such restrictions. Similarly, we note the need for additional research to more fully characterize the prevalence and costs of deaths attributable to harvesting in order to fully understand the cost of pollution and the benefits of any program seeking to address air quality.

While we find some significant effects on mortality, the analysis would be better suited to examining a more sensitive measure of human health, such as hospitalizations. Hence obtaining data on hospitalizations in cities where such short term restrictions are active would be very insightful.

Lastly, our study is unable to disentangle the effects of behavioral responses to pollution alerts from the direct health impacts of lower pollution levels. While the separate characterization of these distinct effects is clearly important, the fact that we see some evidence of gross improvement in health outcomes suggests that a joint governmental approach of information and pollution reduction can effectively reduce the negative health impacts of short term incidents of poor air quality. Notwithstanding, we note how important it is to be able to correctly quantify the extent and costs of behavioral responses to pollution separately from the health impacts of improved air quality, and we hope to address these issues more directly in future research.

3.6 Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in the American Journal of Agricultural Economics, September 2014, aau081. Mullins, Jamie and Prashant Bharadwaj, 2014. The dissertation author was the primary investigator and author of this paper.

3.7 Tables and Figures

Table 3.1: Balance Table

		Means			t-test for Equal Means	
Lag	Variable	Treated	Control	Percent Bias	t-stat	p-value
1	PM_{10}	133.42	136.87	-8.00	-0.32	0.75
2	PM_{10}	129.99	136.77	-15.30	-0.62	0.53
3	PM_{10}	93.83	93.58	0.50	0.03	0.98
4	PM_{10}	85.18	86.70	-3.70	-0.19	0.85
5	PM_{10}	90.22	93.67	-8.20	-0.37	0.71
0	$Temperature$	52.23	51.69	9.10	0.39	0.70
1	$Temperature$	51.54	51.37	2.80	0.12	0.91
2	$Temperature$	50.76	50.44	4.80	0.19	0.85
3	$Temperature$	50.12	50.08	0.60	0.02	0.98
4	$Temperature$	51.15	51.31	-2.70	-0.12	0.91
5	$Temperature$	51.36	51.91	-8.60	-0.36	0.72
0	$Wind\,Speed$	3.35	3.03	26.00	1.12	0.27
1	$Wind\,Speed$	2.70	2.70	-0.40	-0.02	0.98
2	$Wind\,Speed$	2.63	2.60	2.70	0.14	0.89
3	$Wind\,Speed$	3.51	3.55	-3.40	-0.13	0.90
4	$Wind\,Speed$	3.63	3.76	-10.30	-0.44	0.66
5	$Wind\,Speed$	3.07	3.16	-7.50	-0.34	0.74
0	$Precipitation$	0.01	0.01	-2.10	-0.09	0.93
1	$Precipitation$	0.00	0.00	0.00	0.00	1.00
2	$Precipitation$	0.00	0.00	1.30	0.32	0.75
3	$Precipitation$	0.07	0.07	-2.20	-0.07	0.94
4	$Precipitation$	0.05	0.05	-0.30	-0.01	0.99
5	$Precipitation$	0.03	0.02	12.40	0.67	0.51
Observations		34	100			

Notes: Results are based on the 34 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Pollutant concentration data were collected and maintained by Chile's Ministry of the Environment, and weather data are from the NCDC's Summary of the Day data set. Data from 1989-2008 were used for all statistics above.

Precipitation levels on the day before all Episodes and matched days were, in fact, zero. As precipitation tends to dramatically reduce ambient PM_{10} levels, such measurements are not unreasonable.

Table 3.2: Results: Directly Comparing Mean Differences

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Resp. Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-24.07*** (7.17)	-2.41 (2.48)	0.12 (0.95)	-0.66 (0.57)
Difference from Day -1 to Day 1	-35.55*** (10.30)	-5.95 (4.43)	-0.95 (1.70)	-1.59 (1.03)
Difference from Day -1 to Day 2	-28.17*** (10.27)	-7.53 (5.71)	-1.95 (2.46)	-2.021 (1.39)
Difference from Day -1 to Day 3	-27.34** (10.64)	-10.38 (7.33)	-4.27 (3.16)	-2.63 (1.78)
Difference from Day -1 to Day 4	-26.47** (10.91)	-13.84 (9.31)	-4.96 (3.77)	-3.37 (2.23)
Difference from Day -1 to Day 5	-34.52*** (11.22)	-15.24 (10.91)	-6.44 (4.46)	-3.64 (2.64)
Pre-Episode Daily Mean	106.527	64.641	10.971	14.162
Observations	134	119	101	119
Treatment	34	34	34	34
Control	100	85	67	85

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 34 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. Reported pre-Episode means are for the 5 days preceding Episodes which met the common support criteria. Data from 1989-2008 were used for the PM_{10} estimates above. Data constraints limit the mortality analysis and the analysis of respiratory deaths to 1992-2008 and 1994-2008 respectively. Standard errors computed from propensity score matching methods subject to issues as discussed in Caliendo and Kopeinig (2008), bootstrapped standard errors are not presented due to the findings of Abadie and Imbens (2008).

Table 3.3: Main Results: Difference-in-Differences Matching & Regression Results

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Resp. Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-22.527*** (4.99)	-3.611 (2.48)	-0.269 (0.98)	-0.980* (0.58)
Difference from Day -1 to Day 1	-36.647*** (8.01)	-8.281* (4.43)	-1.986 (1.52)	-2.154** (1.06)
Difference from Day -1 to Day 2	-27.898*** (8.15)	-10.664* (5.46)	-2.512 (2.30)	-2.837** (1.36)
Difference from Day -1 to Day 3	-27.258*** (7.29)	-14.756** (7.06)	-4.628 (2.92)	-3.799** (1.75)
Difference from Day -1 to Day 4	-28.237*** (6.84)	-19.436** (8.95)	-5.154 (3.66)	-4.854** (2.18)
Difference from Day -1 to Day 5	-36.164*** (7.39)	-21.824** (10.32)	-7.767* (4.49)	-5.389** (2.53)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Pre-Episode Daily Mean	106.527	64.641	10.971	14.162
Observations	134	119	101	119
Treatment	34	34	34	34
Control	100	85	67	85

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 34 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the treated day.

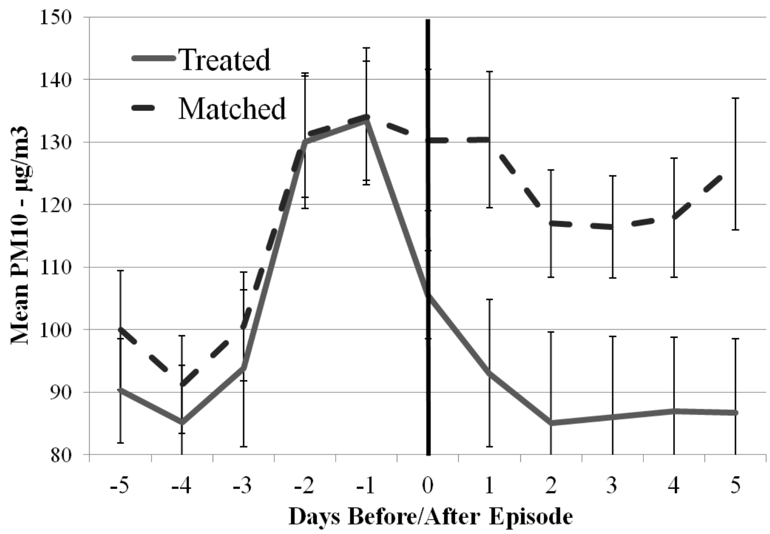
Table 3.4: Effects on Deaths by Cause of Death: All Ages & for those Over 64 Years of Age

	All Ages					Deaths Age Over 64				
	Cum. All Deaths	Cum. Resp. Deaths	Cum. Circ. Deaths	Cum. Cancer Deaths	Cum. Accidental Deaths	Cum. All Deaths	Cum. Resp. Deaths	Cum. Circ. Deaths	Cum. Cancer Deaths	Cum. Accidental Deaths
Difference from Day before to Day of Episode	-7.730***	-0.439	-0.957	-2.105	-0.87	-3.611	-0.269	0.519	-1.467	0.464
	(2.87)	(1.15)	(2.19)	(1.38)	(0.78)	(2.48)	(0.98)	(1.75)	(1.45)	(0.44)
Difference from Day -1 to Day 1	-14.978***	-1.413	-0.572	-4.003	-1.661	-8.281*	-1.986	1.471	-3.42	0.802
	(5.21)	(1.94)	(3.39)	(2.69)	(1.39)	(4.43)	(1.52)	(2.79)	(2.47)	(0.72)
Difference from Day -1 to Day 2	-20.566***	-1.435	-1.597	-4.347	-1.201	-10.664*	-2.512	0.861	-4.524	1.473
	(6.56)	(2.89)	(4.94)	(3.91)	(2.03)	(5.46)	(2.30)	(3.96)	(3.45)	(1.05)
Difference from Day -1 to Day 3	-25.261***	-2.335	-1.861	-5.875	-2.193	-14.756**	-4.628	0.609	-5.771	2.105
	(8.61)	(3.70)	(6.19)	(5.02)	(2.55)	(7.06)	(2.92)	(4.89)	(4.53)	(1.30)
Difference from Day -1 to Day 4	-32.087***	-2.647	-4.625	-8.41	-3.214	-19.436**	-5.154	-1.079	-8.254	2.565*
	(10.83)	(4.62)	(7.61)	(6.08)	(3.01)	(8.95)	(3.66)	(6.16)	(5.57)	(1.50)
Difference from Day -1 to Day 5	-37.418***	-5.082	-5.104	-9.947	-3.543	-21.824**	-7.767*	0.117	-9.972	3.034*
	(12.82)	(5.71)	(9.14)	(7.41)	(3.44)	(10.32)	(4.49)	(7.39)	(6.62)	(1.78)
Pre-Ep. Daily Mean	94.23	12.72	27.62	20.41	4.44	64.64	10.97	22.36	13.28	1.08
N	119	101	101	101	101	119	101	101	101	101
Treatment	34	34	34	34	34	34	34	34	34	34
Control	85	67	67	67	67	85	67	67	67	67

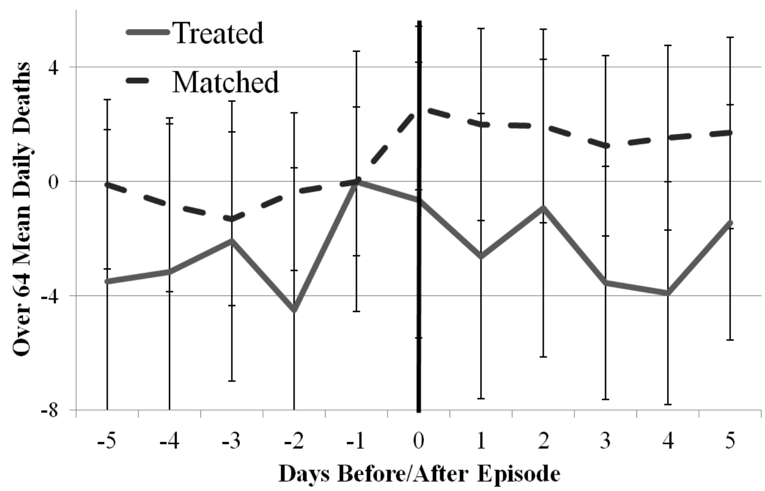
*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 34 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach.



Panel A: PM_{10} before and after Episode



Panel B: Over 64 deaths before and after Episode
Day -1 set to zero

Figure 3.1: Evolution of PM_{10} and over 64 deaths: treated vs. matched

Notes: Error bars show 95% confidence intervals for the means. The vertical line at day zero represents the Episode treatment. Data on historical Episode announcements are available from the Metropolitan Region Ministry of Health. Data for PM_{10} levels were collected and maintained by the National Ministry of the Environment. Mortality data are available from the Chilean Ministry of Health's Department of Statistics and Health Information.

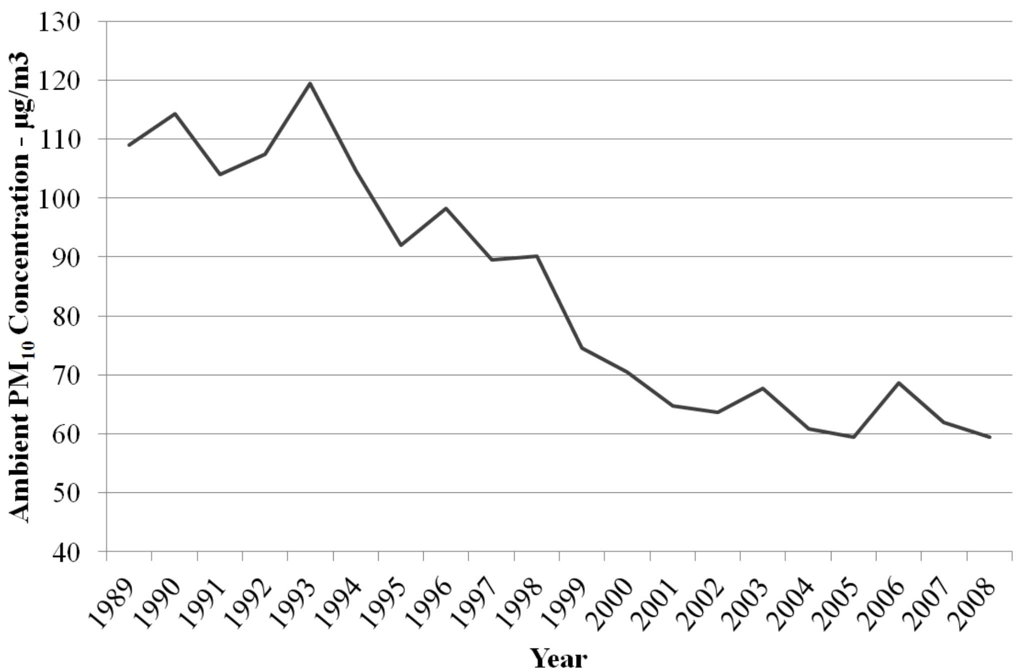


Figure 3.2: Annual mean ambient PM_{10} levels in Santiago

Notes: Data for PM_{10} levels were collected and maintained by the Chilean Ministry of the Environment. Plotted data are annual averages aggregated from daily means of all locations continuously monitored over the study period, specifically the Parque O'Higgins, Las Condes, and La Paz monitoring sites.

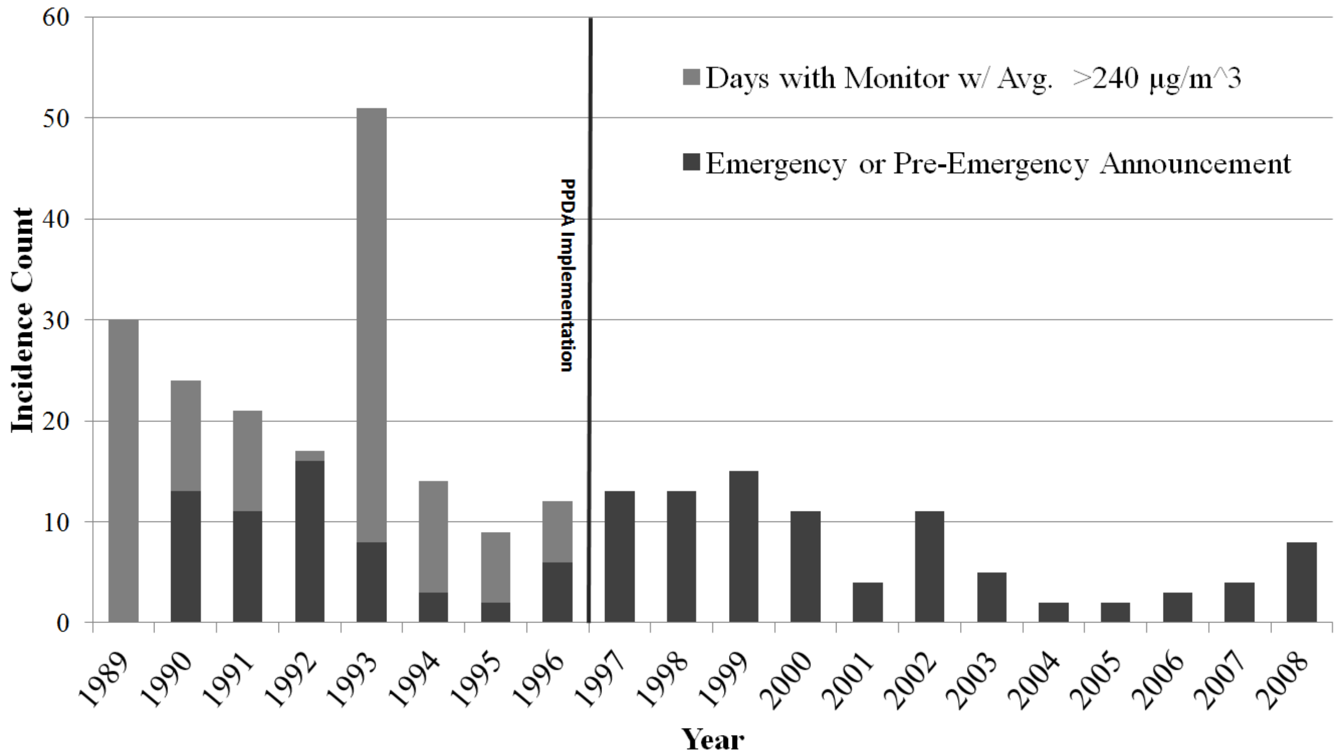


Figure 3.3: Episodes and high-PM_{10} events by year

Notes: Data on historical Episode announcements are available from the Metropolitan Region Ministry of Health. Data for PM_{10} levels were collected and maintained by Chile's Ministry of the Environment.

3.8 Appendix

Table 3.5: Protocols Under the PPDA

Episode Level	Protocols
Seasonal:	
April-August	• Restricted weekday usage of 20% of vehicles without catalytic converters • Implementation of a citywide traffic plan to minimize the effects of the vehicular restrictions
Episodic:	
Alert	• Restricted usage of 40% (weekdays) or 20% (weekends) of vehicles without catalytic converters • Prohibition on the use of uncertified residential wood or biomass heating units
Pre-Emergency	• Restricted usage of 60% (weekdays) or 40% (weekends) of vehicles without catalytic converters • Restricted usage of 20% (all days) of vehicles with catalytic converters • Require operational cessation of plants responsible for 30% of total stationary emissions of particulate matter • Physical Education classes and community sports activities may be suspended by the Ministry of Education • Implementation of more intensive traffic and public transportation plan • Increased enforcement of restrictions on mobile and stationary sources of air pollution • Increased and focused street sweeping and cleaning activities • Increased Metro service schedule implemented • Prohibition of the use of wood or other biomass for residential heating

Table 3.5 - Protocols Under the PPDA - Continued

Emergency	• Restricted usage of 80% (weekday) or 60% (weekend) of vehicles without catalytic converters • Restricted usage of 40% (all days) of vehicles with catalytic converters • Require operational cessation of stationary emissions sources (i.e., plants) responsible for 50% of total stationary emissions • Physical Education classes and community sports activities may be suspended by the Ministry of Education • Implementation of more intensive traffic and public transportation plan • Increased enforcement of vehicle usage restrictions • Increased and focused street sweeping and cleaning activities • Increased Metro service schedule implemented • Prohibition of the use of wood or other biomass for residential heating

Notes: Protocols were adjusted and updated periodically throughout the study period. Those described in this table were in place in 1999. This paper is primarily concerned with Pre-Emergency Episodes because these were both fairly common and involved the shutdown of stationary emissions sources (i.e., plants).

Table 3.6: Example Episode Progression

Day:	Time	Protocol/Action
Day Prior to Episode	8pm	Chilean government announces an Episode for the following day. The decision is based on weather and pollution-level forecasts from a Cassmassi Model ["a multivariate regression tool that weight(s) tendencies on PM_{10} concentrations and 24 h forecasts of five discrete meteorological categories associated with synoptical and subsynoptical features linked to atmospheric stability (i.e., PMCA, Meteorological Potential of Atmospheric Pollution index)"] of the Santiago area and also takes into account the recommendations "made by experienced air quality forecasters" and political considerations (de Grange and Troncoso, 2011; Saide et al., 2011).
Day of Episode	12:01am to 12:00am	Stationary industrial or institutional sources of particulate matter emissions must cease all emitting operations if the particulate matter concentration of their emissions (measured annually by 3rd party assessors and reported to the government) is above the threshold for Pre-Emergency or Emergency Episodes in the current year. Prior to 2001, these thresholds were set such that the sources with the highest emissions concentrations of particulate matter were shut down until 30% or 50% (for Pre-Emergencies and Emergencies respectively) of total daily emissions was restricted for the day. Beginning in 2001, the thresholds were fixed at $32mg/m^3$ and $28mg/m^3$ (for Pre-Emergencies and Emergencies respectively).
	7:30am to 9:00pm	Vehicle restrictions and road closures go into effect. These restrictions changed over time, but have always been based on the last digit of each vehicle's license plate number, and whether or not the vehicle is equipped with a catalytic converter (which is indicated by a government issued green sticker). Traffic police are in charge of enforcing the driving restrictions. During 2008, de Grange and Troncoso (2011) found that Pre-Emergency traffic restrictions (20% of Catalytic Converted vehicles and 60% of vehicles without catalytic converters) led to 5.5% reductions in traffic flows, or 176,000 fewer trips.
	Additional restrictions and protocols are in effect for different periods during the day depending on the level of the Episode. These include increased street sweeping (dust is a major contributor to ambient PM_{10} levels), metro service, and traffic enforcement activities. Additionally, the use of biomass combustion for residential heating is restricted. See table 3.5 for an outline of all restrictions and protocols.	
Day after Episode	If another Episode is not announced, no restrictions remain in place.	

Table 3.7: Table of Means

	1989-1996	1997-2008
Daily Mean PM_{10} ($\mu g/m^3$)	106.15	69.28
Daily Max PM_{10} ($\mu g/m^3$)	128.92	81.23
Temperature ($^\circ F$)	57.85	58.41
Max Daily Temp ($^\circ F$)	74.70	74.32
Min Daily Temp ($^\circ F$)	44.31	45.56
Dew Point ($^\circ F$)	46.17	45.18
Atmospheric Pressure (mb)	959.41	961.45
Visibility ($miles$)	5.04	4.88
Wind Speed ($knots$)	4.72	4.80
Max Wind Speed ($knots$)	12.13	11.65
Wind Gust ($knots$)	25.05	22.36
Daily Precipitation ($inches$)	0.03	0.03
Share of Days w/ Rain	0.12	0.12
Share of Days w/ Fog	0.23	0.22
Share of Days w/ Hail	0.00	0.00
Daily Deaths	77.15	85.79
Daily Deaths Over 64	47.70	57.45
Daily Deaths Under 4	3.40	2.24
Daily Deaths Under 1	3.99	2.62
Daily Respiratory Deaths	6.12	9.22
Daily Respiratory Deaths Over 64	4.74	7.76
Daily Respiratory Deaths Under 1	0.26	0.15
Daily Circulatory Deaths	13.05	23.92
Population	5,496,505	6,342,665
Population Over 64	343,922	469,945
Population Under 4	574,685	519,071
Avg. # Emergency Episodes per Year	0.75	0.17
Avg. # Pre-Emergency Episodes per Year	6.63	7.42

Notes: Population figures are averages across years, and Episode counts are counts per year. All other statistics are daily means, averaged to the annual level, and then the period level.

Table 3.8: Robustness: Effects on Deaths by Age

	Cumulative Deaths			
	All Ages	Deaths Below 1 Year Old	Deaths Below 4 Years Old	Deaths Over 64 Years Old
Difference from Day before to Day of Episode	-7.730***	0.133	-0.261	-3.611
	(2.87)	(0.65)	(0.55)	(2.48)
Difference from Day -1 to Day 1	-14.978***	0.631	-0.108	-8.281*
	(5.21)	(1.16)	(0.89)	(4.43)
Difference from Day -1 to Day 2	-20.566***	0.417	-0.555	-10.664*
	(6.56)	(1.75)	(1.42)	(5.46)
Difference from Day -1 to Day 3	-25.261***	1.036	-0.229	-14.756**
	(8.61)	(2.28)	(1.80)	(7.06)
Difference from Day -1 to Day 4	-32.087***	1.883	0.074	-19.436**
	(10.83)	(2.92)	(2.37)	(8.95)
Difference from Day -1 to Day 5	-37.418***	1.457	-0.786	-21.824**
	(12.82)	(3.46)	(2.86)	(10.32)
Pre-Episode Daily Mean	94.93	2.52	2.94	64.64
N	119	119	119	119
Treatment	34	34	34	34
Control	85	85	85	85

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 34 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the day of treatment. Weather data were taken from the NCDC's Summary of the Day data set. Cumulative mortality estimates are based on data from 1992-2008 due to mortality data availability and changes to the PPDA that were implemented in 2009.

Table 3.9: Robustness: Omitting 1997 and 1998 from the Sample

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-18.881*** (5.50)	-3.611 (2.88)	-0.378 (1.14)	-0.944 (0.65)
Difference from Day -1 to Day 1	-34.039*** (8.66)	-8.773* (4.68)	-1.69 (1.71)	-2.181** (1.08)
Difference from Day -1 to Day 2	-34.385*** (9.39)	-11.219** (5.46)	-2.665 (2.58)	-2.938** (1.34)
Difference from Day -1 to Day 3	-37.649*** (8.57)	-15.352** (7.15)	-4.935 (3.33)	-3.874** (1.76)
Difference from Day -1 to Day 4	-35.940*** (7.14)	-20.038** (8.83)	-6.655 (4.04)	-4.901** (2.13)
Difference from Day -1 to Day 5	-42.475*** (8.42)	-22.725** (10.35)	-9.820* (5.05)	-5.536** (2.51)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	114	104	87	104
Treatment	28	28	28	28
Control	86	76	59	76

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 28 post-PPDA Episodes which occur after 1998, meet our separation criteria, and satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the treated day. Pollutant concentration data were collected and maintained by Chile's Ministry of the Environment, and weather data are from the NCDC's Summary of the Day data set. Data from 1989-2008 were used for the PM_{10} estimates above. Data constraints limit the mortality analysis and the analysis of respiratory deaths to 1992-2008 and 1994-2008 respectively. The period of analysis ends in 2008 due to changes to the PPDA that were implemented in 2009.

Table 3.10: Robustness: Using All Monitors

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-23.195***	-2.351	0.32	-0.608
	(5.15)	(2.63)	(1.19)	(0.61)
Difference from Day -1 to Day 1	-37.307***	-5.234	-1.218	-1.331
	(8.55)	(4.63)	(1.73)	(1.07)
Difference from Day -1 to Day 2	-34.547***	-6.069	-2.748	-1.55
	(9.27)	(5.42)	(2.41)	(1.32)
Difference from Day -1 to Day 3	-22.008***	-9.234	-4.67	-2.198
	(8.04)	(6.98)	(2.93)	(1.69)
Difference from Day -1 to Day 4	-37.200***	-12.741	-4.895	-2.915
	(8.54)	(8.63)	(3.79)	(2.08)
Difference from Day -1 to Day 5	-38.452***	-13.653	-6.331	-3.061
	(9.07)	(9.94)	(4.56)	(2.42)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	140	126	91	126
Treatment	33	33	33	33
Control	107	93	58	93

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 33 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of observations from all in-service monitoring stations on a given day. These include monitors at: Parque O'Higgins, Las Condes, La Paz, Providencia, Cerrillos, El Bosque, La Florida, Cerro Navia, Quilicura, and Pudahuel. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the day of treatment. Data from 1989-2008 were used for the PM_{10} estimates above. Data constraints limit the mortality analysis and the analysis of respiratory deaths to 1992-2008 and 1994-2008 respectively. The period of analysis ends in 2008 due to changes to the PPDA that were implemented in 2009.

Table 3.11: Robustness: Rich Set of Covariates in Propensity Score

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-20.866**	-2.953	-0.673	-0.766
	(7.94)	(3.44)	(1.69)	(0.88)
Difference from Day -1 to Day 1	-47.365***	-10.653	-3.94	-2.683
	(10.69)	(6.64)	(2.65)	(1.69)
Difference from Day -1 to Day 2	-37.107***	-12.601	-4.788	-3.291
	(8.28)	(7.99)	(3.89)	(2.08)
Difference from Day -1 to Day 3	-18.026**	-14.626	-5.554	-3.816
	(8.16)	(10.00)	(4.95)	(2.61)
Difference from Day -1 to Day 4	-43.935***	-19.408	-7.022	-4.979
	(10.55)	(12.98)	(6.20)	(3.36)
Difference from Day -1 to Day 5	-56.179***	-21.041	-7.844	-5.436
	(11.13)	(14.76)	(7.17)	(3.84)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	64	62	53	62
Treatment	19	19	19	19
Control	45	43	34	43

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Propensity Score is generated using a Logit regression of a post-PPDA Episode on 5 days of lags of: daily mean PM_{10}, daily max PM_{10}, wind speed, precipitation, temperature, max temperature, min temperature, atmospheric pressure, dew point, and the square and cube of temperature. Of the 35 post-PPDA Episode which met our separation criteria, only the 18 which also satisfy the common support restrictions of the Propensity Score Matching approach are included in this analysis. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the day of treatment. Data from 1989-2008 were used for the PM_{10} estimates above.

Table 3.12: Robustness: Including Daily Mortality in Propensity Score and Matching

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-30.305*** (6.44)	0.08 (2.36)	0.639 (0.91)	-0.01 (0.55)
Difference from Day -1 to Day 1	-42.907*** (8.98)	-0.79 (4.36)	-0.453 (1.66)	-0.181 (1.03)
Difference from Day -1 to Day 2	-38.180*** (8.18)	-1.57 (5.15)	-1.776 (2.39)	-0.411 (1.29)
Difference from Day -1 to Day 3	-30.301*** (8.12)	-3.228 (6.56)	-3.503 (3.12)	-0.723 (1.63)
Difference from Day -1 to Day 4	-32.449*** (8.36)	-4.931 (8.13)	-4.179 (3.71)	-1.072 (2.00)
Difference from Day -1 to Day 5	-37.300*** (9.16)	-4.112 (9.48)	-5.129 (4.64)	-0.816 (2.35)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	107	107	87	107
Treatment	31	31	31	31
Control	76	76	56	76

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 31 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the treated day. Data from 1989-2008 were used for the PM_{10} estimates above. Data constraints limit the mortality analysis and the analysis of respiratory deaths to 1992-2008 and 1994-2008 respectively. The period of analysis ends in 2008 due to changes to the PPDA that were implemented in 2009.

Table 3.13: Robustness: Alternative Matching Methods

	Mean PM_{10}				Cumulative Deaths Age¿64			
	NN5	NN10	.05 Caliper	Kernel	NN5	NN10	.05 Caliper	Kernel
Difference from Day -1 to Day of Episode	-22.527***	-21.387***	-22.291***	-22.945***	-3.611	-2.248	-2.514	-2.482
	(4.99)	(4.35)	(3.81)	(3.80)	(2.48)	(2.34)	(2.10)	(2.08)
Difference from Day -1 to Day 1	-36.647***	-33.704***	-37.562***	-36.548***	-8.281*	-5.698	-5.869	-5.789
	(8.01)	(7.07)	(5.73)	(5.68)	(4.43)	(4.10)	(3.82)	(3.75)
Difference from Day -1 to Day 2	-27.898***	-28.193***	-34.369***	-34.201***	-10.664*	-7.525	-7.227	-7.161*
	(8.15)	(7.03)	(5.69)	(5.76)	(5.46)	(4.96)	(4.41)	(4.29)
Difference from Day -1 to Day 3	-27.258***	-26.666***	-30.090***	-29.323***	-14.756**	-10.577*	-9.564*	-9.351*
	(7.29)	(6.31)	(5.85)	(5.80)	(7.06)	(6.38)	(5.61)	(5.43)
Difference from Day -1 to Day 4	-28.237***	-33.537***	-35.911***	-34.192***	-19.436**	-14.167*	-12.556*	-12.330*
	(6.84)	(6.27)	(6.02)	(5.96)	(8.95)	(8.26)	(7.24)	(7.03)
Difference from Day -1 to Day 5	-36.164***	-37.653***	-42.823***	-43.182***	-21.824**	-15.473	-13.134	-13.026
	(7.39)	(6.44)	(5.84)	(5.93)	(10.32)	(9.61)	(8.40)	(8.17)
N	134	181	825	828	119	156	612	615
Treatment	34	34	30	30	34	34	30	30
Control	100	147	795	798	85	122	582	585

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10

Eicker-White Standard Errors in Parenthesis ( )

Notes: For ease of comparison, columns 1 & 5 contain the original results from table 3.3 using Nearest 5 Neighbors (NN5) matching method. Columns 2 & 6 contain the results from matching using Nearest 10 Neighbors (NN10), columns 3 & 7 contain matching results using a Caliper matching approach with a 0.05 radius, and columns 4 & 8 contain the results of an Epanechnikov Kernel matching procedure with bandwidth=0.06. Treatments are defined as post-PPDA Episodes meeting our separation criteria, and common support conditions are enforced on all procedures. All regressions include controls discussed previously.

Table 3.14: Robustness: No Separation Criteria Enforced

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-14.233*** (4.39)	-2.165 (1.54)	0.125 (0.67)	-0.556 (0.38)
Difference from Day -1 to Day 1	-21.188*** (5.44)	-3.944 (2.99)	-0.737 (1.24)	-0.967 (0.74)
Difference from Day -1 to Day 2	-22.896*** (6.24)	-6.546 (4.15)	-2.125 (1.90)	-1.581 (1.04)
Difference from Day -1 to Day 3	-19.384*** (5.93)	-8.358 (5.39)	-3.124 (2.51)	-1.946 (1.36)
Difference from Day -1 to Day 4	-28.522*** (5.70)	-11.323* (6.63)	-3.03 (2.98)	-2.588 (1.67)
Difference from Day -1 to Day 5	-39.125*** (6.70)	-13.146* (7.93)	-3.289 (3.52)	-2.935 (1.99)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	278	265	216	265
Treatment	85	85	85	85
Control	193	180	131	180

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on 85 of the 91 post-PPDA Episodes which satisfy the common support restrictions of the Propensity Score Matching approach. No separation criteria is imposed. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of observations from the selected, in-service monitoring stations on a given day. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the day of treatment. Pollutant concentration data were collected and maintained by Chile's Ministry of the Environment, and weather data are from the NCDC's Summary of the Day data set. Data from 1989-2008 were used for the PM_{10} estimates above. Data constraints limit the mortality analysis and the analysis of respiratory deaths to 1992-2008 and 1994-2008 respectively. The period of analysis ends in 2008 due to changes to the PPDA that were implemented in 2009.

Table 3.15: Robustness: Propensity Score Function Weighted Regression Results

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-24.698***	-1.323	-0.387	-0.39
	(4.49)	(2.48)	(0.82)	(0.56)
Difference from Day -1 to Day 1	-35.681***	-4.862	-1.571	-1.285
	(6.62)	(4.49)	(1.42)	(1.00)
Difference from Day -1 to Day 2	-31.588***	-6.728	-2.344	-1.786
	(7.12)	(5.13)	(1.93)	(1.20)
Difference from Day -1 to Day 3	-28.885***	-9.255	-4.097*	-2.401
	(7.47)	(6.44)	(2.47)	(1.50)
Difference from Day -1 to Day 4	-32.960***	-13.149	-4.119	-3.318*
	(6.55)	(8.58)	(2.96)	(1.98)
Difference from Day -1 to Day 5	-38.371***	-14.047	-5.855	-3.515
	(6.15)	(9.86)	(3.58)	(2.31)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	836	623	415	623
Treatment	34	34	34	34
Control	802	589	381	589

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: Results are based on the 34 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. Results are weighted by the function of Propensity Scores put forth by Robins and Ritov (1997), outlined in Imbens (2004), and implemented in Hirano and Imbens (2001). PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. All regressions include controls for temperature, wind, and precipitation on each of the 5 days prior to treatment, and for the month and day-of-week of the day of treatment. Data from 1989-2008 were used for the PM_{10} estimates above. Data constraints limit the mortality analysis and the analysis of respiratory deaths to 1992-2008 and 1994-2008 respectively. The period of analysis ends in 2008 due to changes to the PPDA that were implemented in 2009.

Table 3.16: Robustness: Excluding pre-PPDA Episodes from Control

	Mean PM_{10}	Cum. Deaths Age Over 64	Cum. Over 64 Respiratory Deaths	Cum. Over 64 Death Rate
Difference from Day before to Day of Episode	-20.725*** (5.50)	-2.942 (2.69)	-0.061 (1.00)	-0.8 (0.64)
Difference from Day -1 to Day 1	-38.641*** (7.16)	-7.228 (4.70)	-1.15 (1.62)	-1.919* (1.11)
Difference from Day -1 to Day 2	-38.926*** (7.70)	-10.504* (5.39)	-2.781 (2.29)	-2.823** (1.32)
Difference from Day -1 to Day 3	-30.603*** (7.14)	-13.361* (6.84)	-5.062* (2.81)	-3.505** (1.66)
Difference from Day -1 to Day 4	-36.396*** (7.40)	-17.475** (8.75)	-5.947* (3.51)	-4.534** (2.11)
Difference from Day -1 to Day 5	-46.167*** (8.53)	-19.037* (10.32)	-7.862* (4.26)	-4.929* (2.53)
5-Lags Weather Controls	Yes	Yes	Yes	Yes
DOW Dummies	Yes	Yes	Yes	Yes
Month Dummies	Yes	Yes	Yes	Yes
Observations	123	113	97	113
Treatment	30	30	30	30
Control	93	83	67	83

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: In our headline results, 4 of the days which are matched from the pre-PPDA period were days on which Episodes were announced. Although including such days in our analysis provides conservative estimates of the effects of Episode announcement, we rerun the analysis excluding all Episode days from the matching pool. Presented results are based on the 30 of 35 post-PPDA Episodes meeting our separation criteria, which also satisfy the common support restrictions of the Propensity Score Matching approach. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Death rate statistics are measured in number of deaths per 100,000 residents of the sub-population of interest, and are cumulative from the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day.

Table 3.17: Robustness: Adding in Controls

	Mean PM_{10}				Cumulative Deaths Age Over 64			
	Mean Comparison	+ Weather Controls	+ DOW Controls	+ Month Controls	Mean Comparison	+ Weather Controls	+ DOW Controls	+ Month Controls
Difference from Day -1 to Day of Episode	-24.072***	-23.441***	-22.491***	-22.527***	-2.406	-3.044	-3.216	-3.611
	(7.17)	(5.60)	(4.80)	(4.99)	(2.48)	(2.36)	(2.40)	(2.48)
Difference from Day -1 to Day 1	-35.548***	-36.936***	-36.812***	-36.647***	-5.953	-6.646	-7.425*	-8.281*
	(10.30)	(8.39)	(7.99)	(8.01)	(4.43)	(4.31)	(4.37)	(4.43)
Difference from Day -1 to Day 2	-28.171***	-28.644***	-28.415***	-27.898***	-7.529	-8.176	-9.435*	-10.664*
	(10.27)	(8.00)	(8.07)	(8.15)	(5.71)	(5.52)	(5.45)	(5.46)
Difference from Day -1 to Day 3	-27.336**	-28.440***	-28.702***	-27.258***	-10.382	-11.464	-13.121*	-14.756**
	(10.64)	(7.28)	(7.23)	(7.29)	(7.33)	(7.01)	(7.10)	(7.06)
Difference from Day -1 to Day 4	-26.468**	-29.103***	-29.443***	-28.237***	-13.835	-15.430*	-17.414*	-19.436**
	(10.91)	(6.71)	(6.60)	(6.84)	(9.31)	(8.97)	(8.93)	(8.95)
Difference from Day -1 to Day 5	-34.515***	-36.995***	-36.878***	-36.164***	-15.241	-17.562*	-19.339*	-21.824**
	(11.22)	(7.32)	(7.36)	(7.39)	(10.91)	(10.28)	(10.29)	(10.32)
5-Lags Weather Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
DOW Dummies	No	No	Yes	Yes	No	No	Yes	Yes
Month Dummies	No	No	No	Yes	No	No	No	Yes
N	134	134	134	134	119	119	119	119
Treatment	34	34	34	34	34	34	34	34
Control	100	100	100	100	85	85	85	85

*** - significant at 0.01; ** - significant at 0.05; * - significant at 0.10
Eicker-White Standard Errors in Parenthesis ( )

Notes: For ease of comparison, columns 1 & 5 contain the original results from table 3.2 while columns 4 & 8 repeat the mainline results from table 3.3. The results reported in columns 2 & 6 add weather controls to the straight comparison of means, and columns 3 & 7 further add controls for the day of week on which the treatment occurred. Finally, month controls were added in columns 4 & 8 to duplicate our mainline specification. Treatments are defined as post-PPDA Episodes meeting our separation criteria, and common support conditions are enforced on all procedures. PM_{10} values are in terms of concentrations measured in $\mu g/m^3$. Death statistics are reported in total number of deaths, and are cumulative beginning at the time of treatment. Calculations are based on city-wide averages of the daily means of PM_{10} observations from the selected, in-service monitoring stations on a given day. Weather controls include temperature, wind, and precipitation on each of the 5 days prior to treatment. Pollutant concentration data were collected and maintained by Chile's Ministry of the Environment, and weather data are from the NCDC's Summary of the Day data set. Data from 1989-2008 were used for the PM_{10} estimates above, while data from 1992-2008 were used for mortality estimates. The period of analysis ends in 2008 due to changes to the PPDA that were implemented in 2009.

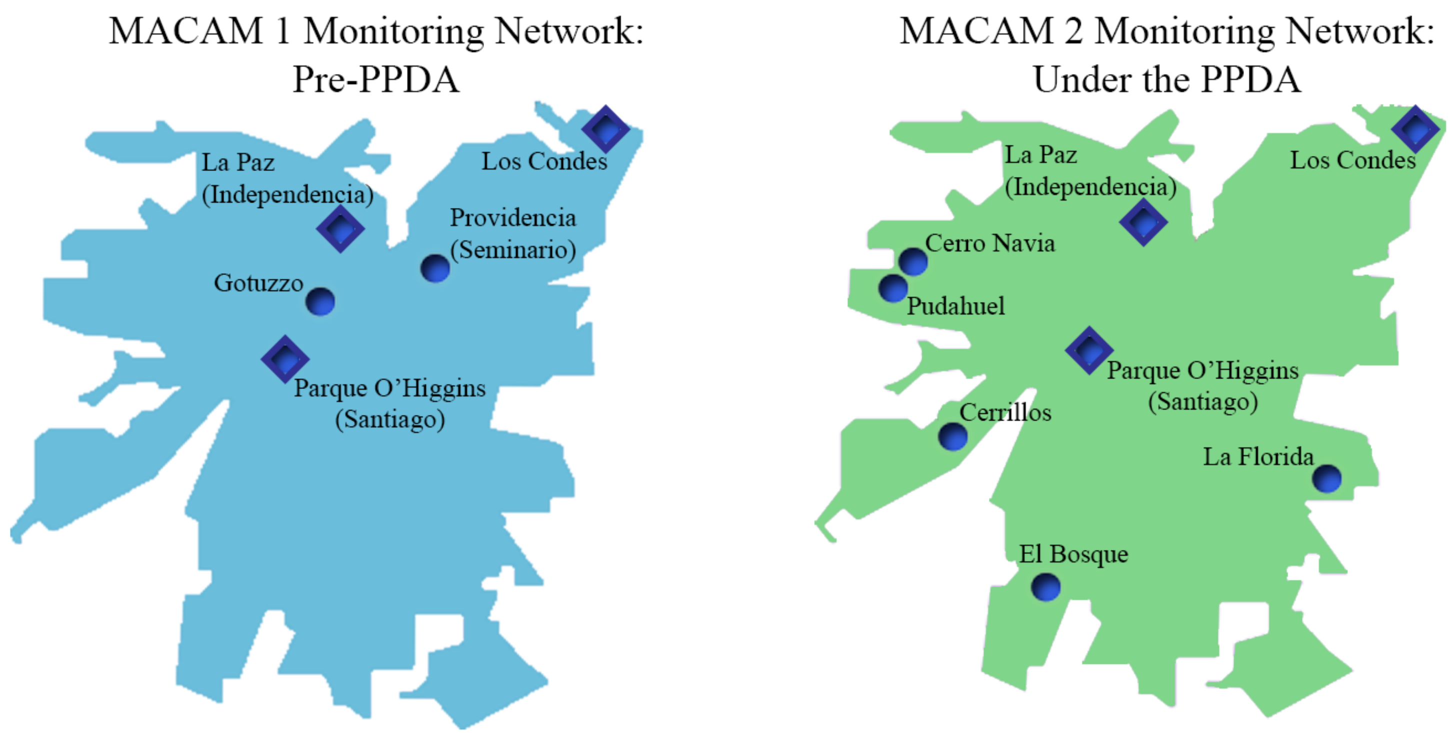


Figure 3.4: Maps of pollution monitoring stations in Santiago

Notes: Maps are adaptions of those produced in Chilean Ministry of the Environment (2007). Diamonds mark the monitors used in the headline results.

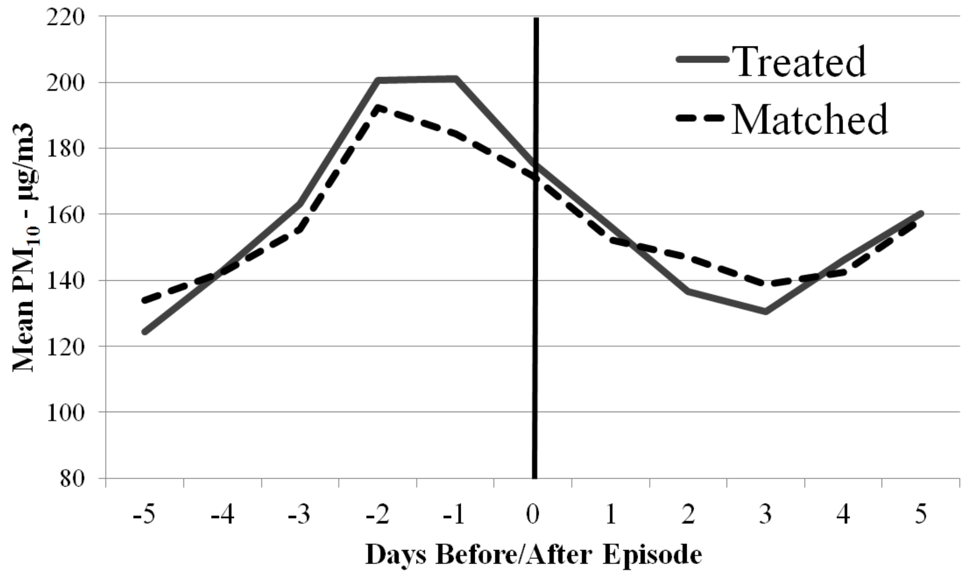


Figure 3.5: PM_{10} before and after Episode: pre-PPDA Episodes vs. matched

Notes: Data on historical Episode announcements are available from the Metropolitan Region Ministry of Health. Data for PM_{10} levels are from the National Ministry of the Environment.

Bibliography

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

_ **and Guido W. Imbens**, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 2005, *74* (1), 235–267.

_ **and** _ , "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 2008, *76* (6), 1537–1557.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham, "Management and Shocks to Worker Productivity: Evidence from Air Pollution Exposure in an Indian Garment Factory," *Working Paper*, 2014.

Agency, U.S. Environmental Protection, "National Air Quality and Emissions Trends Report, 1994," 1994, pp. 2–6 – 2–8.

Anderson, Jonathan O., Josef G. Thundiyil, and Andrew Stolbach, "Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health," *Journal of Medical Toxicology*, 2012, *8* (2), 166–175.

Auffhammer, Maximilian and Ryan Kellogg, "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality," *The American Economic Review*, 2011, *101* (6), 2687–2722.

Babcock, Philip, "Real Costs of Nominal Grade Inflation? New Evidence from Student Course Evaluations," *Economic Inquiry*, 2010, *48* (4), 983–996.

Bandyopadhyay, Sushenjit and John Horowitz, "Do Plants Overcomply with Water Pollution Regulations? The Role of Discharge Variability," *Topics in Economic Analysis & Policy*, 2006, *6* (1).

Bascom, Rebecca, Philip A Bromberg, Daniel A Costa, Robert Devlin, Douglas W Dockery, Mark W Frampton, William Lambert,

Jonathan M Samet, Frank E Speizer, and Mark Utell, "Health effects of outdoor air pollution," *American Journal of Respiratory and Critical Care Medicine*, 1996, *153* (1), 3–50.

Becker, William E and Sherwin Rosen, "The Learning Effect of Assessment and Evaluation in High School," *Economics of Education Review*, 1992, *11* (2), 107–118.

Betts, Julian R and Jeff Grogger, "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings," *Economics of Education Review*, 2003, *22* (4), 343–352.

Binkley, Helen M, Joseph Beckett, Douglas J Casa, Douglas M Kleiner, and Paul E Plummer, "National Athletic Trainers' Association Position Statement: Exertional Heat Illnesses," *Journal of Athletic Training*, 2002, *37* (3), 329.

Brauer, Michael, Jim Blair, and Sverre Vedal, "Effect of Ambient Ozone Exposure on Lung Function in Farm Workers.," *American Journal of Respiratory and Critical Care Medicine*, 1996, *154* (4), 981–987.

Brunekreef, Bert, Gerard Hoek, Oscar Breugelmans, and Marcel Leentvaar, "Respiratory Effects of Low-Level Photochemical Air Pollution in Amateur Cyclists," *American Journal of Respiratory and Critical Care Medicine*, 1994, *150* (4), 962–966.

Caiazzo, Fabio, Akshay Ashok, Ian A Waitz, Steve HL Yim, and Steven RH Barrett, "Air pollution and early deaths in the United States. Part I: Quantifying the impact of major sectors in 2005," *Atmospheric Environment*, 2013, *79*, 198–208.

Caliendo, Marco and Sabine Kopeinig, "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, 2008, *22* (1), 31–72.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2).

Carson, Richard T., Michael B. Conaway, W. Michael Hanemann, Raymond J. Kopp, K. Martin, RC Mitchell, S. Presser, and PA Ruud, "Valuing Winter Visibility Improvement in the Grand Canyon," in "Association of Environmental and Resource Economists Meeting. New Orleans" 1992.

Chang, Tom, Joshua S. Graff Zivin, Tal Gross, and Matthew J. Neidell, "Particulate Pollution and the Productivity of Pear Packers," Working Paper 19944, National Bureau of Economic Research February 2014.

Chay, Kenneth, Carlos Dobkin, and Michael Greenstone, "The Clean Air Act of 1970 and adult mortality," *Journal of Risk and Uncertainty*, 2003, *27* (3), 279–300.

Chay, Kenneth Y. and Michael Greenstone, "Air Quality, Infant Mortality, and the Clean Air Act of 1970," *Working paper, National Bureau of Economic Research*, 2003.

_ **and** _ , "Does Air Quality Matter? Evidence From the Housing Market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Chilean Ministry of Health, "Supreme Decree 32," 1990.

Chilean Ministry of the Environment, "Gestion de La Calidad Del Aire: Varias Decadas de Esfuerzo," 2007.

Clark, Derek J and Christian Riis, "Competition Over More Than One Prize," *American Economic Review*, *88* (1), 276–289.

Coria, Jessica, "Environmental Policy, Fuel Prices and the Switching to Natural Gas in Santiago, Chile," *Working Paper*, 2007.

_ , "Environmental Policy, Fuel Prices and the Switching to Natural Gas in Santiago, Chile," *Ecological Economics*, 2009, *68* (11), 2877–2884.

Currie, Janet and Matthew Neidell, "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?," *The Quarterly Journal of Economics*, 2005, *120* (3), 1003–1030.

_ **, Eric A. Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G. Rivkin**, "Does Pollution Increase School Absences?," *The Review of Economics and Statistics*, 2009, *91* (4), 682–694.

_ **, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell**, "What Do We Know About Short-and Long-Term Effects of Early-Life Exposure to Pollution?," *Annual Review of Resource Economics*, 2014, *6* (1), 217–247.

Cutter, W. Bowman and Matthew Neidell, "Voluntary Information Programs and Environmental Regulation: Evidence from 'Spare the Air'," *Journal of Environmental Economics and Management*, 2009, *58* (3), 253–265.

Czibor, Eszter, Sander Onderstal, Randolph Sloof, and Mirjam van Praag, "Does Relative Grading Help Male Students? Evidence from a Field Experiment in the Classroom," *Working Paper*, 2014.

Davis, Lucas W., "The Effect of Driving Restrictions on Air Quality in Mexico City," *Journal of Political Economy*, 2008, *116* (1), 38–81.

de Grange, Louis and Rodrigo Troncoso, "Impacts of Vehicle Restrictions on Urban Transport Flows: The Case of Santiago, Chile," *Transport Policy*, 2011, *18* (6), 862–869.

Deschenes, Olivier, Michael Greenstone, and Joseph S. Shapiro, "Defensive Investments and the Demand for Air Quality: Evidence From the NOx Budget Program and Ozone Reductions," *Working paper, National Bureau of Economic Research*, 2012.

Devlin, Robert B, William F McDonnell, Richard Mann, Susanne Becker, Dennis E House, Dina Schreinemachers, and Hillel S Koren, "Exposure of Humans to Ambient Levels of Ozone for 6.6 Hours Causes Cellular and Biochemical Changes in the Lung," *American Journal of Respiratory Cell and Molecular Biology*, 1991, *4* (1), 72–81.

Dimeo, MJ, MG Glenn, MJ Holtzman, JR Sheller, JA Nadel, and HA Boushey, "Threshold concentration of ozone causing an increase in bronchial reactivity in humans and adaptation with repeated exposures.," *The American review of respiratory disease*, 1981, *124* (3), 245–248.

Dubey, Pradeep and John Geanakoplos, "Grading Exams: 100, 99, 98,... or A, B, C?," *Games and Economic Behavior*, 2010, *69* (1), 72–94.

Duffield, Rob, Brian Dawson, and Carmel Goodman, "Energy system contribution to 400-metre and 800-metre track running," *Journal of Sports Sciences*, 2005, *23* (3), 299–307.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-Telling By Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India," *The Quarterly Journal of Economics*, 2013, *1499* (1545), 1499.

Environmental Protection Agency, "International Transport of Air Pollution," *Our Nation's Air - Status and Trends through 2008*, 2010, pp. 43–44.

_ , "Clean Power Plan Proposed Rule," June 2014.

_ , "Method 5: Determination of Particulate Matter Emissions from Stationary Sources."

Farrell, BP, HD Kerr, TJ Kulle, LR Sauder, and JL Young, "Adaptation in human subjects to the effects of inhaled ozone after repeated exposure.," *The American review of respiratory disease*, 1979, *119* (5), 725–730.

Figlio, David N and Maurice E Lucas, "Do High Grading Standards Affect Student Performance?," *Journal of Public Economics*, 2004, *88* (9), 1815–1834.

Fung, Karen, Daniel Krewski, Rick Burnett, and Francesca Dominici, "Testing the Harvesting Hypothesis by Time-Domain Regression Analysis. I: Baseline Analysis," *Journal of Toxicology and Environmental Health, Part A*, 2005, *68* (13-14), 1137–1154.

Geiser, J, M Vogt, R Billeter, C Zuleger, F Belforti, and H Hoppeler, "Training high-living low: changes of aerobic performance and muscle structure with training at simulated altitude," *International journal of sports medicine*, 2001, *22* (8), 579–585.

Gerrity, Timothy R, Richard A Weaver, J Berntsen, Dennis E House, and John J O'Neil, "Extrathoracic and intrathoracic removal of O 3 in tidal-breathing humans," *Journal Applied Physiology*, 1988, *65* (1), 393–400.

Ghanem, Dalia and Junjie Zhang, "'Effortless Perfection': Do Chinese Cities Manipulate Air Pollution Data?," *Journal of Environmental Economics and Management*, 2014, *68* (2), 203–225.

Ghio, Andrew J, Chong Kim, and Robert B Devlin, "Concentrated ambient air particles induce mild pulmonary inflammation in healthy human volunteers," *American Journal of Respiratory and Critical Care Medicine*, 2000, *162* (3), 981–988.

Giles, Luisa V. and Michael S. Koehle, "The Health Effects of Exercising in Air Pollution," *Sports Medicine*, 2014, *44* (2), 223–249.

Gold, Diane R, Augusto Litonjua, Joel Schwartz, Eric Lovett, Andrew Larson, Bruce Nearing, George Allen, Monique Verrier, Rebecca Cherry, and Richard Verrier, "Ambient pollution and heart rate variability," *Circulation*, 2000, *101* (11), 1267–1273.

Gómez, Federico P, Veronica M Amado, Josep Roca, Antoni Torres, Josep M Nicolas, Robert Rodriguez-Roisin, and Joan A Barberà, "Effect of nitric oxide inhalation on gas exchange in acute severe pneumonia," *Respiratory physiology & neurobiology*, 2013, *187* (2), 157–163.

Graff Zivin, Joshua and Matthew Neidell, "Days of Haze: Environmental Information Disclosure and Intertemporal Avoidance Behavior," *Journal of Environmental Economics and Management*, 2009, *58* (2), 119–128.

_ **and** _ , "The Impact of Pollution On Worker Productivity," *American Economic Review*, 2012, *102* (7), 3652–3673.

Gramsch, Ernesto, Francisco Cereceda-Balic, Pedro Oyola, and Dietrich Von Baer, "Examination of Pollution Trends in Santiago de Chile with Cluster Analysis of PM10 and Ozone Data," *Atmospheric Environment*, 2006, *40* (28), 5464–5475.

Green, Jerry R and Nancy L Stokey, "A Comparison of Tournaments and Contracts," *The Journal of Political Economy*, 1983, *91* (3), 349–364.

Greenstone, Michael, Aparna Krishnan, Rohini Pande, Nicholas Ryan, and Anant Sudarshan, "Improving Human Health Through a Market-Friendly Emissions Scheme," in "Seminar on Global Environment and Disaster Management: Law and Society organized by the Supreme Court of India, July," Vol. 23 2011.

Griffiths, Mark JD and Timothy W Evans, "Inhaled Nitric Oxide Therapy in Adults," *New England Journal of Medicine*, 2005, *353* (25), 2683–2695.

Hackney, Jack D, William S Linn, Sarunas K Karuza, Ramon D Buckley, David C Law, David V Bates, Milan Hazucha, L David Pengelly, and Frances Sil Verman, "Effects of Ozone Exposure in Canadians and Southern Californians: Evidence for Adaptation?," *Archives of Environmental Health: An International Journal*, 1977, *32* (3), 110–116.

Hancock, Peter A, "A dynamic model of stress and sustained attention," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1989, *31* (5), 519–537.

Harris, Gardiner, "Beijing's Bad Air Would Be Step Up for Smoggy Delhi," *New York Times*, January 25, 2014.

Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith, “The Economics and Econometrics of Active Labor Market Programs,” *Handbook of Labor Economics*, 1999, *3*, 1865–2097.

Hesterberg, Thomas W, William B Bunn, Roger O McClellan, Ali K Hamade, Christopher M Long, and Peter A Valberg, “Critical review of the human data on short-term nitrogen dioxide (NO2) exposures: evidence for NO2 no-effect levels,” *Critical reviews in toxicology*, 2009, *39* (9), 743–781.

Heyes, Anthony, “Implementing Environmental Regulation: Enforcement and Compliance,” *Journal of Regulatory Economics*, 2000, *17* (2), 107–129.

Hill, Archibald Vivian, “The physiological basis of athletic records.,” *The Lancet*, 1925, *206* (5323), 481–486.

Hirano, Keisuke and Guido W. Imbens, “Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data On Right Heart Catheterization,” *Health Services and Outcomes Research Methodology*, 2001, *2* (3), 259–278.

Holguín, Fernando, Marta M Téllez-Rojo, Mauricio Hernández, Marlene Cortez, Judith C Chow, John G Watson, David Mannino, and Isabelle Romieu, “Air pollution and heart rate variability among the elderly in Mexico City,” *Epidemiology*, 2003, *14* (5), 521–527.

Horvath, SM, JA Gliner, and LJ Folinsbee, “Adaptation to ozone: duration of effect.,” *The American review of respiratory disease*, 1981, *123* (5), 496–499.

Huanxin, Zhao, “Beijing Pollution Fight 'Greater Than for Olympics',” *China Daily*, March 13, 2013.

Imbens, Guido W., “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review,” *Review of Economics and Statistics*, 2004, *86* (1), 4–29.

Jerrett, Michael, Richard T Burnett, C Arden Pope, Kazuhiko Ito, George Thurston, Daniel Krewski, Yuanli Shi, Eugenia Calle, and Michael Thun, “Long-term ozone exposure and mortality,” *New England Journal of Medicine*, 2009, *360* (11), 1085–1095.

Kang, Suk, “A Formal Model of School Reward Systems,” *Incentives, Learning and Employability, edited by John Bishop, Columbus Ohio: National Center for Research in Vocational Education*, 1985, pp. 956–971.

Kehrl, HR, LM Vincent, RJ Kowalsky, DH Horstman, JJ O'Neil, WH McCartney, and PA Bromberg, "Ozone Exposure Increases Respiratory Epithelial Permeability in Humans.," *The American Review of Respiratory Disease*, 1987, *135* (5), 1124–1128.

Kinsella, John P, Gary R Cutter, William F Walsh, Dale R Gerstmann, Carl L Bose, Claudia Hart, Kris C Sekar, Richard L Auten, Vinod K Bhutani, Jeffrey S Gerdes et al., "Early inhaled nitric oxide therapy in premature newborns with respiratory failure," *New England Journal of Medicine*, 2006, *355* (4), 354–364.

Koren, Hillel S, Robert B Devlin, Delores E Graham, Richard Mann, Maria P McGee, Donald H Horstman, Walter J Kozumbo, Susanne Becker, Dennis E House, William F McDonnell et al., "Ozone-Induced Inflammation in the Lower Airways of Human Subjects," *American Review of Respiratory Disease*, 1989, *139* (2), 407–415.

Korrick, Susan A, Lucas M Neas, Douglas W Dockery, Diane R Gold, George A Allen, L Bruce Hill, Kenneth D Kimball, Bernard A Rosner, and Frank E Speizer, "Effects of ozone and other pollutants on the pulmonary function of adult hikers.," *Environmental health perspectives*, 1998, *106* (2), 93.

Langrish, Jeremy P, Magnus Lundbäck, Stefan Barath, Stefan Söderberg, Nicholas L Mills, David E Newby, Thomas Sandström, and Anders Blomberg, "Exposure to nitrogen dioxide is not associated with vascular dysfunction in man," *Inhalation toxicology*, 2010, *22* (3), 192–198.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth, "The Impact of Air Pollution on Cognitive Performance and Human Capital Formation," *Working Paper: Royal Holloway University of London*, 2012.

Lazear, Edward P, "Speeding, Terrorism, and Teaching to the Test," *The Quarterly Journal of Economics*, 2006, pp. 1029–1061.

_ **and Sherwin Rosen**, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 1981, pp. 841–864.

Lee, David S, "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Levy, Ilan, Cristian Mihele, Gang Lu, Julie Narayan, and Jeffrey R Brook, "Evaluating Multipollutant Exposure and Urban Air Quality: Pollutant Interrelationships, Neighborhood Variability, and Nitrogen Dioxide as a Proxy Pollutant," *Environmental Health Perspectives*, 2014, *122* (1), 65.

Likens, Gene E., Charles T. Driscoll, and Donald C. Buso, "Long-Term Effects of Acid Rain: Response and Recovery of a Forest Ecosystem," *Science*, 1996, *272* (5259), 244–246.

Lin, C.Y. Cynthia, Wei Zhang, and Victoria I. Umanskaya, "On the Design of Driving Restrictions: Theory and Empirical Evidence," *Working paper, University of California at Davis*, 2013.

Linder, J, D Herren, Ch Monn, and HU Wanner, "Effect of Ozone on Physical Performance," *Schweizerische Zeitschrift fur Sportmedizin*, 1988, *36* (1), 5.

Linn, William S, Joseph C Solomon, Stanley C Trim, Charles E Spier, Deborah A Shamoo, Theodore G Venet, Edward L Avol, and Jack D Hackney, "Effects of exposure to 4 ppm nitrogen dioxide in healthy and asthmatic volunteers," *Archives of Environmental Health: An International Journal*, 1985, *40* (4), 234–239.

Lippmann, Morton, *Environmental Toxicants: Human Exposures and Their Health Effects*, 3rd ed., New York: John Wiley & Sons,

_ , Terry Gordon, and Lung Chi Chen, "Effects of Subchronic Exposures to Concentrated Ambient Particles in Mice: IX. Integral Assessment and Human Health Implications of Subchronic Exposures of Mice to CAPs," *Inhalation Toxicology*, 2005, *17* (4-5), 255–261.

Lleras-Muney, Adriana, "The Needs of the Army Using Compulsory Relocation in the Military to Estimate the Effect of Air Pollutants on Children's Health," *Journal of Human Resources*, 2010, *45* (3), 549–590.

Logan, WPD, "Mortality in the London Fog Incident, 1952," *The Lancet*, 1953, *261* (6755), 336–338.

Luechinger, Simon, "Valuing Air Quality Using the Life Satisfaction Approach," *The Economic Journal*, 2009, *119* (536), 482–515.

McDonnell, William F, Donald H Horstman, Louis J Raggio, James A Green et al., "The Respiratory Responses of Subjects with

Allergic Rhinitis to Ozone Exposure and Their Relationship to Nonspecific Airway Reactivity," *Toxicology and Industrial Health*, 1987, *3* (4), 507–517.

McGrath, JJ, "Biological Plausibility for Carbon Monoxide as a Copollutant in PM epidemiologic studies.," *Inhalation Toxicology*, 2000, *12*, 91.

Montero, Juan-Pablo, Jose Miguel Sanchez, and Ricardo Katz, "A Market-Based Environmental Policy Experiment in Chile," *Journal of Law and Economics*, 2002, *45* (1), 267–287.

Moretti, Enrico and Matthew Neidell, "Pollution, Health, and Avoidance Behavior," *Journal of Human Resources*, 2011, *46* (1), 154–175.

Mullins, Jamie and Prashant Bharadwaj, "Effects of Short-Term Measures to Curb Air Pollution: Evidence from Santiago, Chile," *American Journal of Agricultural Economics*, 2014.

Nalebuff, Barry J and Joseph E Stiglitz, "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *The Bell Journal of Economics*, 1983, pp. 21–43.

Neidell, Matthew, "Information, Avoidance Behavior, and Health the Effect of Ozone on Asthma Hospitalizations," *Journal of Human Resources*, 2009, *44* (2), 450–478.

Omori, Takashi, Go Fujimoto, Isao Yoshimura, Hiroshi Nitta, and Masaji Ono, "Effects of Particulate Matter on Daily Mortality in 13 Japanese Cities.," *Journal of Epidemiology/Japan Epidemiological Association*, 2003, *13* (6), 314–322.

Ostro, Bart D., "The Effects of Air Pollution on Work Loss and Morbidity," *Journal of Environmental Economics and Management*, 1983, *10* (4), 371–382.

Palacios, Milagros and Carlos Chávez, "Determinants of Compliance in the Emissions Compensation Program in Santiago, Chile," *Environment and Development Economics*, 2005, *10* (04), 453–483.

Pande, Rohini, Deanna Ford, Nick Ryan, and Anant Sudarshan, "Industrial Pollution, Regulation and Growth: Governance Challenges and Innovations," *Working Paper. The 2012 San Servolo Workshop on Grand Challenges of Sustainability, Sustainability Science Program*, 2012.

Paredes, Valentina, "Grading System and Student Effort," *U.C. Berkeley Job Market Paper*, 2012.

Peng, Roger D., Francesca Dominici, and Thomas A. Louis, "Model Choice in Time Series Studies of Air Pollution and Mortality," *Journal of the Royal Statistical Society: Series A*, 2006, *169* (2), 179–203.

Pope, C Arden, "Respiratory Disease Associated with Community Air Pollution and a Steel Mill, Utah Valley.," *American Journal of Public Health*, 1989, *79* (5), 623–628.

Prendez, Margarita, Gerardo Alvarado, and Italo Serey, *"Some Guidelines to Improve the Air Quality Management of Santiago, Chile." In N. Mazzeo, ed. Air Quality Monitoring, Assessment and Management*, Rijeka, Croatia: InTech, 2011.

Roberts, Jesse D, Jeffrey R Fineman, Frederick C Morin, Philip W Shaul, Stephen Rimar, Michael D Schreiber, Richard A Polin, Maurice S Zwass, Michael M Zayek, Ian Gross et al., "Inhaled nitric oxide and persistent pulmonary hypertension of the newborn," *New England Journal of Medicine*, 1997, *336* (9), 605–610.

Robins, James M. and Andrea Rotnitzky, "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 1995, *90* (429), 122–129.

_ and Ya'acov Ritov, "Toward a Curse of Dimensionality Appropriate (Coda) Asymptotic Theory for Semiparametric Models," *Statistics in Medicine*, 1997, *16* (3), 285–319.

_ , Andrea Rotnitzky, and Lue Ping Zhao, "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 1995, *90* (429), 106–121.

Roger, LJ, HR Kehrl, M Hazucha, and DH Horstman, "Bronchoconstriction in Asthmatics Exposed to Sulfur Dioxide During Repeated Exercise," *Journal of Applied Physiology*, 1985, *59* (3), 784–791.

Rosenbaum, Paul R. and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, *70* (1), 41–55.

Rossaint, R, K Lewandowski, and WM Zapol, "Our paper 20 years later: Inhaled nitric oxide for the acute respiratory distress syndrome-discovery, current understanding, and focussed targets of future applications," *Intensive care medicine*, 2014, *40* (11), 1649–1658.

Rossaint, Rolf, Konrad J Falke, Frank Lopez, Klaus Slama, Ulrich Pison, and Warren M Zapol, "Inhaled nitric oxide for the adult respiratory distress syndrome," *New England Journal of Medicine*, 1993, *328* (6), 399–405.

Rutllant, Jose and Rene Garreaud, "Meteorological Air Pollution Potential for Santiago, Chile: Towards an Objective Episode Forecasting," *Environmental Monitoring and Assessment*, 1995, *34* (3), 223–244.

Saide, Pablo E., Gregory R. Carmichael, Scott N. Spak, Laura Gallardo, Axel E. Osses, Marcelo A Mena-Carrasco, and Mariusz Pagowski, "Forecasting Urban PM10 and PM2.5 Pollution Episodes in Very Stable Nocturnal Conditions and Complex Terrain Using WRF-Chem CO Tracer Model," *Atmospheric Environment*, 2011, *45* (16), 2769–2780.

Saltin, Bengt and Philip D Gollnick, "Skeletal Muscle Adaptability: Significance for Metabolism and Performance," *Comprehensive Physiology*, 2011.

Samet, Jonathan M., Francesca Dominici, Frank C. Curriero, Ivan Coursac, and Scott L. Zeger, "Fine Particulate Air Pollution and Mortality in 20 US Cities, 1987–1994," *New England Journal of Medicine*, 2000, *343* (24), 1742–1749.

Sanders, Nicholas J. and Charles F. Stoecker, "Where Have All the Young Men Gone? Using Gender Ratios to Measure Fetal Death Rates," *National Bureau of Economic Research*, 2011.

Schelegle, Edward S and William C Adams, "Reduced Exercise Time in Competitive Simulations Consequent to Low Level Ozone Exposure," *Medicine & Science in Sports & Exercise*, 1986, *18* (4), 408–414.

Schlenker, Wolfram and W. Reed Walker, "Airports, Air Pollution, and Contemporaneous Health," *National Bureau of Economic Research*, 2011.

_ **and W Reed Walker**, "Airports, Air Pollution, and Contemporaneous Health," *Working Paper: University of California, Berkeley*, 2012.

Schofield, Hugh, "Paris car ban imposed after pollution hits high," *BBC News*, March 17, 2014.

Schulz, Richard and Christine Curnow, "Peak Performance and Age Among Auperathletes: Track and Dield, Swimming, Baseball, Tennis, and Golf," *Journal of Gerontology*, 1988, *43* (5), P113–P120.

Selwyn, Beatrice J, Thomas H Stock, DE Jenkins, Dennis J Kotchmar, and Robert S Chapman, "Health Effects of Ambient Ozone Exposure in Vigorously Exercising Adults," Technical Report, Texas Univ., Houston (USA). School of Public Health 1986.

Shavell, Steven, *Economic Analysis of Accident Law*, Harvard University Press, 2009.

Shimshack, Jay P and Michael B Ward, "Enforcement and Over-Compliance," *Journal of Environmental Economics and Management*, 2008, *55* (1), 90–105.

Siegel, Ron, "All-Pay Contests," *Econometrica*, 2009, *77* (1), 71–92.

_ , "Head Starts in Contests," *Working Paper, Northwestern University*, 2011.

Smith, Jeffrey A. and Petra E. Todd, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 2005, *125* (1), 305–353.

Spektor, Dalia M, Morton Lippmann, George D Thurston, Paul J Lioy, Jurij Stecko, George O'Connor, Eric Garshick, Frank E Speizer, and Carl Hayes, "Effects of Ambient Ozone on Respiratory Function in Healthy Adults Exercising Outdoors," *American Review of Respiratory Disease*, 1988, *138* (4), 821–828.

Stevenson, DS, FJ Dentener, MG Schultz, K Ellingsen, TPC Van Noije, O Wild, G Zeng, M Amann, CS Atherton, N Bell et al., "Multimodel Ensemble Simulations of Present-Day and Near-Future Tropospheric Ozone," *Journal of Geophysical Research: Atmospheres*, 2006, *111* (D8).

Teman, Nicholas R, Jeffrey Thomas, Benjamin S Bryner, Carl F Haas, Jonathan W Haft, Pauline K Park, Mark J Lowell, and Lena M Napolitano, "Inhaled Nitric Oxide to Improve Oxygenation for Safe Critical Care Transport of Adults With Severe Hypoxemia," *American Journal of Critical Care*, 2015, *24* (2), 110–117.

Troncoso, Rodrigo, Louis de Grange, and Luis Cifuentes, "Effects of Environmental Alerts and Pre-Emergencies On Pollutant Concentrations in Santiago, Chile," *Atmospheric Environment*, 2012, *61*, 550–557.

Wayne, Walborg S, Paul F Wehrle, and Robert E Carroll, "Oxidant Air Pollution and Athletic Performance," *JAMA*, 1967, *199* (12), 901–904.

Weitzman, Martin L, "Prices vs. Quantities," *The Review of Economic Studies*, 1974, pp. 477–491.

World Health Organization, "Air Quality and Health," September 2011.

Yang, Biao, Neil D Burns, and Chris J Backhouse, "Management of Uncertainty Through Postponement," *International Journal of Production Research*, 2004, *42* (6), 1049–1064.

Zanobetti, Antonella, Joel Schwartz, Evi Samoli, Alexandros Gryparis, Giota Touloumi, Richard Atkinson, Alain Le Tertre, Janos Bobros, Martin Celko, Ayana Goren et al., "The Temporal Pattern of Mortality Responses to Air Pollution: A Multicity Assessment of Mortality Displacement," *Epidemiology*, 2002, *13* (1), 87–93.

Zeger, Scott L., Francesca Dominici, and Jonathan Samet, "Harvesting-Resistant Estimates of Air Pollution Effects on Mortality," *Epidemiology*, 1999, *10* (2), 171–175.

Zweig, Jacqueline S, John C Ham, and Edward L Avol, "Air Pollution and Academic Performance: Evidence from California Schools," *Working Paper: Department of Economics, University of Maryland*, 2009.