Measuring the Long-Term Impact of Environmental Externalities

Steffen Andersen, Danmarks Nationalbank and CEPR

Luigi Guiso, EIEF and CEPR

Federico Marciano, University of Chicago

Paola Sapienza, Northwestern University, NBER, and CEPR

Luigi Zingales, University of Chicago, NBER, and CEPR*

Abstract

We exploit a unique dataset of the Danish population observed over more than four decades to measure the long-term health effects of toxic pollution. Making use of incinerators' openings and closings, we explore the interactions among cumulative exposure, intensity, and latency as determinants of lung cancer incidence. On average, the impact of an incinerator becomes detectable 25-30 years after the opening and persists for 10-15 years after the closing. An average amount of cumulative exposure increases the risk of lung cancer by 8%. At low levels of intensity, cumulative exposure impacts lung cancer with a decade-long latency (up to 15 years). At high levels, latency is much shorter (less than 5 years). Not accounting for latency leads to underestimate the health effects of cumulative exposure to incinerators by 90 to 100%, depending on intensity.

Preliminary and Incomplete

Current Draft: March 2023

JEL Classification: JEL I1, K32, Q5, Q53

Keywords: Pollution; PM2.5 Mortality; Waste Incinerators; Lung Cancer, Latency.

*We thank Jim Poterba and participants to lunch seminars at MIT, University of Chicago, and Northwestern University for very useful comments. We also thank Lorenzo Bruno and Cornelius Arendt Skytte Petersen for the excellent research assistantship. Luigi Zingales gratefully acknowledges financial support from the Stigler Center at the University of Chicago.

Introduction

Ex-ante regulation and ex-post legal liability are key mechanisms in modern economies to curb the negative effects of the pollution generated by agricultural and industrial production. Both these remedies, however, rely on relatively timely estimates of the damages caused by pollution (e.g., Currie et al., 2015). If one is interested in reducing the incidence of chronic diseases like cancer, a major obstacle to this timely estimation is represented by latency, defined as the delay between exposure to the agent causing the disease and the diagnosis. Lung cancer, for instance, is associated with an average latency of 13.6 years (Nadler and Zurbenko, 2014). As the evidence of smokers suggests, both the magnitude of the effect and the length of latency are a function of the lifetime cumulative exposure to a toxic agent and of the intensity of this exposure (Lubin et al., 2007). Given the complexity of the interplay among cumulative exposure, intensity, and latency as well as the mobility of people in and out of polluted areas, we are not aware of any estimate of the latency and hence of the total carcinogenic effects even of the most widespread forms of air pollution like fine particulate matter (PM2.5).

The goal of this paper is to provide such estimates by exploiting a unique dataset of the Danish population observed over more than four decades. This data contains residential address, demographic and financial information as well as health outcomes, such as cancer diagnoses. In addition to this dataset with detailed individual-level information, Denmark is an interesting case-study because of its diffuse use of waste incinerators and its thorough record of the emissions produced by incinerators. Thus, we exploit the openings and closings of these facilities to estimate the long-term impact of exposure to toxic emissions on lung cancer incidence.

Before implementing our empirical strategy based on openings and closings, we describe how it is designed to address the four main challenges that we face. First, exposure to emissions is hard to measure, varying with distance from the sources, weather conditions, height of the chimney, and quantity of pollutants emitted. Second, our dataset does not include smoking habits, known to be an important determinant of lung cancer. Third, locations exposed to incinerators can be characterized by different background levels of pollution; in addition, individuals can sort themselves geographically, based on socioeconomic status, health conditions, and levels of attention they pay to their health. Last but not least, the interaction among cumulative exposure

to pollution, intensity, latency, and cancer diagnoses is complex and hard to estimate empirically.

We solve the first problem by relying on HYSPLIT, an air dispersion model that takes as inputs the coordinates of multiple emission sources, the amount of their emissions, the height of the chimney, and weather conditions to produce an estimate of the concentration of the pollutants of interest. We validate the quality of these predictions with both satellite-based measures of pollution concentration (from the MERRA-2 dataset provided by the NASA) and data from pollution monitors installed by the Danish EPA.

To address the second problem, for the subset of women who became pregnant during the last 20 years, we can observe smoking habits and verify that they are not positively correlated with exposure to incinerators (if any, the correlation is negative) and that the latter is a significant predictor of lung cancer even after adding smoking habits as a control.

As for the third problem, we have the possibility to exploit the openings and the closings of incinerators as exogenous shocks to predict lung cancer. In all our specifications, we can control for location fixed effects that capture any heterogeneity in background pollution. Evolution in exposure to background pollution common to all the citizens in the country can be captured by a set of time fixed effects. We can also control for a number of important individual covariates, like income, wealth, sex, and marital status; these variables *i*) are available in our records, *ii*) can be merged with the health status data, and *iii*) are of administrative origin and thus measured with great accuracy. Observing exposed and unexposed individuals before the events that we want to study, we can rule out the presence of any economically significant differences in the dependent variable and in all our controls. In addition, we document that neither the openings nor the closings of incinerators have any impact on the decision of exposed individuals to move to another area.

As for the fourth problem, with the aim of disentangling the effects of cumulative exposure, intensity, and latency, we can consider the closings of incinerators. More in detail, observing individuals who have been exposed to emissions from an incinerator that is shut down, we can keep cumulative exposure fixed and assess whether excess lung cancer risk persists after the end of exposure due to the presence of latency. In addition, we can see whether the length of latency is affected by the intensity of exposure prior to the closing.

After addressing these four threats to identification, we implement our empirical strategy. The medical literature on cancer development (e.g., Pitot, 1993) suggests that the relationship between exposure to incinerators and lung cancer, if any, should not be contemporaneous, but characterized by a delay. In a difference-in-differences framework, we exploit the openings of incinerators to determine whether this relationship exists and to measure the presence and the length of the delay between the start of exposure and the manifestation of an excess risk of lung cancer for the exposed individuals as compared to the unexposed. We show that, on average, exposed individuals start experiencing an excess risk of lung cancer between 25 and 30 years after the opening. The magnitude of this excess risk is 4 basis points, with the baseline for the relevant population being 35 (hazard ratio of 1.11). The excess risk cannot be explained by age and variables such as income, wealth, sex, and marital status. We also find that for people associated with the highest level of intensity of exposure to the incinerator at the time of its opening, the manifestation of an excess risk of lung cancer is much faster.

By looking at openings, we document the presence of a delay between the start of exposure and the increase in lung cancer risk. This delay may be due to the effect of cumulative exposure (i.e., people need to accumulate significant amounts of cumulative exposure before developing the disease) or latency (i.e., there is a delay between exposure to toxic pollution and the development of lung cancer). Our test based on openings does not allow us to estimate separately the magnitude of these two potential effects, because individuals accumulate additional exposure to polluting sources while being observed during the post-opening period.

To separate cumulative exposure from latency, we focus on the closings of incinerators. We study the health outcomes of residents who have been exposed to an incinerator, after the plant is shut down and the residents do not accumulate exposure anymore. According to our results, exposed individuals keep experiencing an excess risk of lung cancer for up to 15 years after the closing, due to the presence of latency. This excess risk is proportional to the amount of exposure accumulated before the closing and an exposed individual with an average value of cumulative exposure is associated with a hazard ratio of 1.08. When we differentiate the impact for both the intensity of exposure and the time elapsed since the closing, we find that there is a relationship between the two. For the highest category of intensity, excess cancer diagnoses are

concentrated in the five years after the closing. By contrast, for lower levels of intensity, the delayed effect manifests itself after 5 years and up to 15 years. In general, there seems to be an inverse relationship between the intensity of exposure and the length of the latency period.

Our analysis suggests that considering the interactions among cumulative exposure, intensity, and latency is crucial in computing the health damages of pollution as well as the benefits of any environmental regulation that reduces emissions. Not accounting for latency would lead to capture only 4% of the overall number of excess lung cancer cases related to the post-closing period but caused by exposure accumulated before the closing. The magnitude of the underestimation depends on the intensity of exposure and ranges between 90 (for higher exposure) and 100% (for lower exposure). Based on our estimates from the sample of closings, we also run a simulation by considering a plant that pollutes at the intensity currently associated with Amager Bakke, the incinerator located in the center of Copenhagen. Our results suggest that it may take up to 30 years after the opening to detect the effect of such an incinerator on lung cancer incidence. In addition, a difference-in-differences estimator that does not rely on a follow-up of at least 10 years during the post-treatment period may completely fail to capture the effect of an environmental policy reducing emissions from such an incinerator to zero.

Finally, our estimates allow us to assess incinerators as a source of pollution in Denmark and to evaluate the reduction in their impact on health over time, considering the introduction of two emissions caps in 1991 and in 2003. We estimate that the median incinerator active in Denmark in 1990 was generating 5 excess lung cancer cases (with a latency of 5-15 years) per year of activity. In 2004, after the two emissions caps, the same incinerator generated 0.5 excess lung cancer cases (with a latency of 10-15 years) per year of activity. As a result, following the two regulatory interventions, the health consequences of the activity of such an incinerator have decreased by a factor of 10. Still, the long-term impact of its environmental externalities accounts for around 40% of its revenues.

Our paper is related to three strands of research: studies in environmental and health economics about the relationship between exposure to air pollution and health outcomes, epidemiological studies about carcinogenic air pollutants, and the literature on occupational exposure and on smoking habits.

As for the first strand of research, economists have shed light on how air pollution affects health outcomes by exploiting natural experiments with the aim of addressing the sources of endogeneity that characterize exposure (Graff Zivin and Neidell, 2013). Important examples are provided by Chay and Greenstone (2003), Currie and Neidell (2005), Currie and Schmieder (2009), Currie and Walker (2011), Moretti and Neidell (2011), Currie et al. (2015), and Deryugina et al. (2019). The great contribution of this strand of research has been to provide causal estimates of the impact of pollution on health. However, due to the inability to track the residential addresses of many individuals over time, these studies may underestimate the overall effects as they have focused only on contemporaneous health outcomes (mostly infant health), without considering chronic diseases and exploring the (possibly delayed) effects of cumulative exposure to air pollution.

While environmental and health economists have not focused on chronic diseases like cancer, many studies in epidemiology have sought to explore the link between air pollution and different cancer sites. The most studied relationship is the one between PM2.5 and lung cancer. Two authoritative examples are Krewski et al. (2009) and Raaschou-Nielsen et al. (2013)[1]. Both papers document that an increment in the concentration of PM2.5 of $10\mu g/m^3$ is associated with a lung cancer hazard ratio of 1.10, consistent with other studies in the literature. The great contribution of this strand of research has been to shed light on the lung cancer implications of air pollution, which can pose health hazards to every individual, regardless of occupation and lifestyle. The main limitation of these studies is that they do not offer a rigorous analysis of cumulative exposure and latency, using measures of exposure to air pollution that do not account for lifetime exposure history at the individual level[2].

[1]These two papers are cited as the main supporting evidence by the International Agency for Research on Cancer to justify its decision to consider air pollution and, in particular, PM2.5 as a Group 1 carcinogenic substance in 2013; see https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Air-Pollution-And-Cancer-2013

[2]Raaschou-Nielsen et al. (2013) consider the concentration of PM2.5 measured at the residential address of survey participants at the time of enrollment as their explanatory variable of interest. Krewski et al. (2009) adopt the same strategy as Raaschou-Nielsen et al. (2013) in their main analysis and then propose an estimation based on time windows (i.e., in the same regression, they consider as explanatory variables the averages of the concentration of PM2.5 at the residential address computed with reference to the past 1-4, 5-9, 10-14, and 15-19 years); they find that no time window appears to be more important than the others as a predictor of lung cancer, due to the correlation in exposure over time for the same individual. There are even examples of epidemiological studies in this field that completely overlook latency and cumulative exposure by considering the contemporaneous concentration of PM2.5 at the residential address as their main explanatory variable (e.g., Shin et al., 2022, Laden et al., 2006, Hystad et al., 2013, Nafstad et al., 2003, and Lepeule et al., 2012).

Finally, studies on occupational exposure and on smoking habits have thoroughly investigated the roles of cumulative exposure, intensity, and latency to predict different cancer sites (Richardson, 2009 mentions the most relevant references). The advantage of these studies as opposed to epidemiological papers on the cancer implications of air pollution is that they are able to reconstruct the exposure history of the subjects under investigation in much detail. In addition, in their framework, the study of latency can be performed relatively easily, because it is immediate to identify both the start and the end of exposure (e.g., a person starts and quits smoking at specific points in time). In contrast, in the context of air pollution, exposure history for a given individual is typically continuous, making it hard to identify such turning points as the start and the end of exposure, which are critical to the modelling of latency. On the other hand, the main limitation of studies on occupational exposure and on smoking habits is that they work with very selected samples of individuals. As a consequence, their implications are hard to extend to ordinary people who do not have a particular occupation and are not smokers.

Our paper aims to be the first to offer quasi-experimental evidence of the health implications of long-term exposure to air pollution by relying on individual-level data and by modelling the interplay among cumulative exposure, intensity, and latency. The closest studies are four. Chen et al. (2013) and Ebenstein et al. (2017) estimate the impact of long-term exposure to air pollution in China by following a regression-discontinuity design. Anderson (2020) exploits random variation in exposure generated by wind direction to estimate the effect of air pollution from highways on mortality in the US. Barreca et al. (2021) exploit the Acid Rain Program in the US to estimate the effect of a reduction in SO2 emissions on mortality. To the best of our knowledge, these studies represent the only economic papers to explore the link between measures of long-term exposure to air pollution and health. The main limitations that our paper seeks to overcome are two. First, all these studies work with aggregate data at the level of some geographical unit; more in detail, Chen et al. (2013) and Ebenstein et al. (2017) collapse measures of health and pollution at the city level, Anderson (2020) work with census blocks, and Barreca et al. (2021) consider counties; as a consequence of this limitation, no one of these papers is able to reconstruct lifetime exposure history at the individual level, which is essential to the purposes of our analysis. Second, no one of these studies explicitly models latency, except

for Barreca et al. (2021), which studies how the effect of the Acid Rain Program on mortality evolved over a 15-year time since its implementation in 1995.

1 Data

Our data consists of three different datasets: individual-level data on health and economic status, data on incinerators, and information on air quality.

1.1 Individual-Level Data

Our main source for individual-level information is a unique dataset that covers the entire universe of Danish people from 1975 to 2018. For every subject, we can observe the geo-coded residential address (from the Danish Civil Registration System) from 1975 to 2018; the 1975 addresses are associated with the year when the subject of interest first moved to the relevant address, back to 1900.

In order to match individual-level data with environmental data on exposure to incinerators (from the dispersion model) and air quality in general (from the NASA), we need to make use of a grid fitted to a geographical unit. Given that in our data addresses are already matched with parishes and that parishes constitute a small administrative unit in Denmark, we chose to use Danish parishes as the relevant grid. In Denmark, there are 2,161 parishes, with an average of 2,380 people per parish. This matching allows us to assign to each individual a measure of exposure to pollution.

From 1980 to 2018, we can also observe demographic variables (age, sex, and marital status from the Danish Civil Registration System), financial data (income and wealth from the Danish Tax Authority), and health outcomes (cause of death and cancer cases by cancer site from the Cause of Death Registry and the National Patient Registry respectively) for each Danish person. Every subject is also associated with a personal identification number given at birth and a household identification number constructed by Statistics Denmark, which allows us to track individuals and generation links over time. In particular, we compute both wealth and income at the household level, given that we are interested in proxying for overall socioeconomic status

with the aim of predicting health outcomes; both variables are expressed in 2015-DKK.

In all our specifications, we lag income and wealth by two years, given that there is a mechanical relationship between the manifestation of events such as cancer diagnosis and a decline in economic status (Chetty et al., 2016). As a result, we lose two years of data for every individual and the sample for which we have the full set of variables consists of the Danish population aged at least 2 observed between 1982 and 2018 (187,050,733 person-years).

Since our analysis focuses on the cumulative amount of exposure to incinerators, we are forced to drop all the individuals without information on residential address for at least one year and all the individuals who spend at least one year in a Municipality associated with an incinerator for which we are not able to identify opening and closing years (Helsingor, Svejstrup, Skibby, Struer, and Ringsted). This reduces the sample size to 169,820,085 person-years.

1.2 Data on Emissions from Incinerators

We collect data on the 77 incinerators active in Denmark for the period between 1975 and 2018, when individual-level data on residential addresses are available. The first piece of information concerns location, opening and closing (if any) years, and height of the chimney; our sources are Kleis and Dalager (2007) and the local archives of the municipalities associated with the incinerators[3]. We are not able to identify the opening and closing years for five incinerators, located in Helsingor, Svejstrup, Skibby, Struer, and Ringsted. Consequently, we drop these five incinerators and our dataset consists of 72 unique incinerators, with 1,526 incinerator-years observed between 1975 and 2018.

For each incinerator-year, we collect information on annual emissions of the toxic pollutants relevant to our analysis aimed at predicting lung cancer (SO2, PM2.5, dioxins, and heavy metals). Our data sources for annual emissions are the Green Accounts[4] (mandatory environmental reports introduced in 1996 and containing information about annual emissions at the plant level), the Danish Energy Agency[5] (which provides data on the annual amount of waste burnt from 1994 for all the active incinerators), and the CORINAIR Program[6], which provides data on annual

[3] https://arkiv.dk/
[4] https://miljoeoplysninger.mst.dk/PrtrPublicering/Dataservice
[5] https://ens.dk/en
[6] https://www.eea.europa.eu/publications/EMEPCORINAIR/page005.html

sector-specific emissions factors related to incinerators, defined as tonnes of emissions of a given pollutant of interest per tonne of waste burnt in one year, spanning the period from 1990 onward.

The Green Accounts represent our preferred source, because annual emissions are directly measured at the plant level; the reliability of these measurements is ensured by central supervisory authorities assigned to every plant by the Danish Ministry of Environment and Food, which go through each annual report verifying the completeness and quality of information (Amore et al., 2019).

Unfortunately, this source is available only from 1996 and it does not cover all the incinerators, but only the ones above size thresholds established by law[7]. Furthermore, it does not include all the pollutants of interest. While SO2 is consistently reported by all the incinerators subjected to the reporting requirement, dioxins are reported only from 2003 onward, PM2.5 is never reported but only data on PM10 are available, and among the heavy metals only data on lead are complete. Finally, at the moment, we obtained access only to the Green Accounts published in 1996-2001 and in 2007-2018[8]; in the second reporting period, we still need to access some of the reports, which are not public due to a confidentiality clause introduced in 2010.

In addition to "actual emissions" as reported in the Green Accounts, we can compute "imputed emissions" as the product between the CORINAIR emissions factors and the annual amount of waste burnt. While the amount of waste burnt is available only since 1994, the advantages of this imputation method are twofold. First, we can observe all the incinerator-years of interest from 1994 onward, given that the Danish Energy Agency provides the annual amount of waste burnt for all the plants included in our dataset. Second, we can have emissions for all the relevant pollutants, because the CORINAIR Program covers PM2.5, dioxins, and heavy metals, in addition to SO2.

[7]The first size threshold for incinerators was introduced in 2002, when the Green Accounts Act was modified to exempt small entities from the reporting requirement (Thy, 2003 and Holgaard and Jørgensen, 2005)

[8]In 2015, the Green Accounts were abolished and replaced with the PRTR reports established by the EU regulation; for the purposes of our analysis, the difference between the two types of environmental report is irrelevant.

1.3 Data on Air Quality

Finally, we also collect data on air quality in Denmark. We download the MERRA-2 dataset provided by the NASA[9], which allows us to observe the monthly concentration of SO2 and of the precursors of PM2.5 at the spatial horizontal resolution of 0.5^{o}x0.625^{o} from 1980 to the present. We compute PM2.5 concentration from its precursors[10], we collapse monthly observations of PM2.5 and SO2 concentrations at the annual frequency, and we match the grid of the NASA with the shapefile of the Danish parishes by means of spatial averages, so that each person-year can be assigned a level of SO2 and PM2.5 concentration based on his/her residential address (which is matched with a parish).

In addition to the NASA dataset, we download annual average concentrations of SO2 and PM2.5 registered by the network of measuring stations managed by the University of Aarhus as part of the Air Quality Monitoring Program of the Danish EPA[11].

1.4 Data on Smoking Habits for Pregnant Women

For all the Danish pregnant women between 1999 and 2018 (1,199,750 pregnant-woman-years), we observe a dummy variable ("Smoker") that takes the value of 1 if a pregnant woman is classified as a current smoker. Upon one of the first visits under pregnancy, the family doctor in collaboration with the pregnant woman fills out a pregnancy journal, identifying the due date and recovering background health information relevant to the pregnancy. Smoking behavior is part of such information. We use this data source to test whether smoking habits and exposure to pollution from incinerators are positively correlated; in order to attain this aim, we match pregnant-woman-years with our master dataset, where we draw information on residential address and our set of controls; when performing the matching, we lose 31 observations for which income and wealth are not available.

[9] https://disc.gsfc.nasa.gov/datasets/M2TMNXAER_5.12.4/summary

[10] PM2.5 = DUSMASS25 + OCSMASS+ BCSMASS + SSSMASS25 + SO4SMASS* (132.14/96.06); see https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/FAQ/

[11] https://www2.dmu.dk/1_Viden/2_miljoe-tilstand/3_luft/4_maalinger/5_database/hentdata_en.asp

2 History of Incinerators and Regulation

2.1 History of Incinerators in Denmark

Denmark pioneered the use of incinerators as a way of reducing the amount of land devoted to landfills and the risk of contamination of the aquifer, which represents an existential threat in a flat country. The first incineration plant was built in Frederiksberg in 1903[12]. Yet, until the mid-1960s, there were only three incinerators. After the 1973 oil crisis, to reduce energy dependence, Denmark started to build more and more incinerators. By 1984, Denmark had 40 working incinerators (Weibust, 2005).

The first concerns about the health effects of incinerators started to surface in the mid-1980s. In 1984, the Danish EPA wrote a report that documented the quantity of dioxins emitted by incinerators (Kleis and Dalager, 2007). As a result, the worst emitters came under strong pressure to shut down. Since dioxins tend to be emitted more by small plants, which do not operate at sufficiently high temperatures, the closed plants tend to be small. During the period for which we have all the controls (1982-2018), 42 different plants were closed, with the highest concentration in the period 1984-1993 (17 closings).

At the same time, new incinerators were built and old ones were rebuilt according to the most forefront of green technology. The most striking example is Amager Bakke, the incineration plant built in the center of Copenhagen in 2016, which produces zero emissions of dioxins and can be also used as a ski slope[13]. Of the new incinerators, 15 were built in a municipality that did not have incinerators before, while 25 in the same place as old incinerators.

As a result of all these openings and closings, the number of incinerators declined to 24 in 2018. Nevertheless, some incinerators continue to create problems. In 2019, Norfors I/S received an injunction from the Danish EPA because it had been violating the limit on dioxin emissions from incineration plants since 2014[14]. In 2021, Ea Energy Analyses (a consulting company which closely collaborates with the Danish Government) published a report, in which

[12]Even though the first incinerator was activated in 1903, we compute exposure to pollution from incinerators starting from 1950. Indeed, it is after World War 2 that the incineration industry started representing a non-negligible source of toxic pollution in Denmark (Kleis and Dalager, 2007).

[13]https://www.copenhill.dk/en

[14]https://zerowasteeurope.eu/2019/07/the-case-of-the-danish-norfors-plant/

it recommends that 9 out of the 24 Danish incinerators should be shut down by 2030[15].

2.2 Environmental Regulation in Denmark

The Danish EPA was created in 1974. The first executive order to cap toxic emissions from incinerators, however, was issued only in 1991[16]. It imposed limits on emissions of particulate matter (PM10), arsenic, heavy metals, but not dioxins.

After the 1984 report, the Danish EPA recommended the introduction of legal limits on dioxins' emissions, but this recommendation was ignored. In the 1990s, the EU started working on a directive to impose such a limit, but the process was slowed down by intense lobbying (Kleis and Dalager, 2007).

Only in 2000 was the EU directive on dioxins emissions caps approved. Denmark implemented this directive in 2003 with a second executive order, which reduced also the limits for emissions of particulate matter and heavy metals[17].

2.3 Green Accounts

In 1996, the Danish EPA mandated the most energy-intensive plants (including all incinerators) to disclose the consumption of resources, emissions, and amounts of waste in separate accounts, called "Green Accounts". These environmental reports also contained information on the environmental strategy pursued by firms and were freely accessible to the public.

Such a form of disclosure faced enormous opposition from the Danish industry (Thy, 2003 and Holgaard and Jørgensen, 2005). Thus, in subsequent years a series of reforms ended up reducing the set of plants required to report. In 2002, stringent size thresholds were introduced, exempting from the reporting requirements the smallest entities (and also the smallest incinerators) by number of employees and capacity. In addition, in 2010 the disclosure of the Green Accounts to the public was made optional. Finally, in 2015 the Green Accounts were replaced with the Pollutant Release and Transfer Register (PRTR) reports mandated by the EU[18].

[15] chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.kl.dk/media/25918/kapacitetstilpasningsplan-for-affald-analyserapport.pdf

[16] https://www.retsinformation.dk/eli/lta/1991/10

[17] https://www.retsinformation.dk/eli/lta/2003/162

[18] https://miljoeoplysninger.mst.dk/PrtrPublicering/OmGroenneRegnskaber

Both the Green Accounts and the PRTR reports have been stored over the years by the Danish EPA, from which we obtained access to the data.

3 Medical Mechanism and Threats to Identification

3.1 Medical Mechanism

The medical consensus (e.g., Pitot, 1993) has it that cancer development consists of three main stages. The first, cancer initiation, is triggered by a DNA mutation that transforms a normal cell into a cancer cell. The second, cancer promotion, consists of the replication of the initial cancer cell. The third, cancer progression, is the continuation of cancer promotion and involves the transformation of a benign tumor into a neoplasm and into malignancy[19]. Note that the human body has its own mechanisms to repair DNA damage and to react to the replication of the first cancer cell, slowing down the process. For these reasons, decade-long latency periods can separate the first mutation in the DNA of the initial cancer cell from the detection of the disease by means of a diagnosis. For example, Nadler and Zurbenko (2014) document that 35 out of 44 cancer sites are associated with more than 10 years of latency. In particular, lung cancer has a latency of 13.6 years

Carcinogenic substances can be divided into three categories: cancer initiators, cancer promoters, and strong carcinogens. The first group of substances contributes to cancer initiation by determining the initial DNA mutation that gives rise to cancer development. The second group of substances contributes to cancer promotion by facilitating and accelerating the replication of pre-existing cancer cells. The third group of substances is able to contribute to both stages. To give some examples, tobacco smoke is a strong carcinogen, but its cancer-promotion properties are thought to prevail over the cancer-initiation effects (Weisburger, 1990). PM2.5 has been recently found to be a cancer promoter (Swanton, 2023).

Given that incinerators are heavy emitters of PM2.5 (Krause and Smith, 2006), this classification of carcinogenic substances is very relevant for our key hypotheses and identification strategy. If the substance of interest acts as a cancer promoter, cumulative exposure is essential

[19] https://www.cancerquest.org/cancer-biology/cancer-development

to the acceleration of a latent cancer development process. In addition, if the exposure to the cancer promoter ends (e.g., a smoker decides to quit, or the emission source is shut down), the excess risk of cancer should vanish some years after the end of exposure, when the human body has time to destroy the existing cancer cells, whose replication is not facilitated by the action of an external agent anymore. Due to the roles played by latency and cumulative exposure, we expect the effect of the start of exposure to an incinerator on lung cancer incidence to materialize with a delay. Due to the fact that the impact of cancer promoters can be reversed if exposure ends, we also expect that the effect of cumulative exposure may vanish some years after the closing of a plant, depending on the length of the latency period.

3.2 Threats to Identification

3.2.1 Measuring Exposure Correctly

In estimating the health effects of pollution, we face several challenges. The first one is purely technical. The health effects vary with exposure and different individuals are exposed in different ways, depending upon the distance from the polluting sources, the quantity of pollutants emitted, the height of the chimney of the polluters, and the weather conditions prevailing at the time of the emissions.

We solve this problem by relying on an air dispersion model that takes as inputs the coordinates of multiple emission sources, the quantity of their emissions, the height of the chimney, and the weather conditions to produce an estimate of the concentration of the pollutants of interest in correspondence to different geographical areas. We discuss this model, its application, and its validation in Section 4.

3.2.2 Unobserved Heterogeneity

The second challenge comes from the fact that people are not randomly assigned to cumulative doses of pollution, thus there can be systematic difference in unobservable characteristics that predict lung cancer. The most worrisome is smoking. Smoking habits are the single most important predictor of lung cancer[20] and, while we do control for financial demographic

[20] https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm

information that can predict smoking habits very well, we do not have information on smoking behavior for the whole sample. If smoking habits are correlated with exposure to emissions from incinerators (e.g., smokers tend to live in municipalities with incinerators or closer to them), our results could be spurious.

Fortunately, we do have information on smoking habits for the women who became pregnant in the period 1999 to 2018, since they had to discuss their smoking habits with their family doctors during one of the first visits under pregnancy. Using these data, we conduct two robustness tests (see Subsection 6.4).

First, we test whether there is any correlation between smoking habits and exposure to incinerators' pollution among pregnant women. We find that if anything smoking women are less likely to locate near incinerators. Thus, if anything the unobserved heterogeneity is likely to bias our results downward.

Second, we check that our results carry through if we restrict the analysis to women who have become pregnant at least once between 1999 and 2018, after we control for smoking habits. They do.

3.2.3 Endogeneity of Location

A more severe version of the problem of non-random treatment is that exposure is not only not random, but it is also the result of conscious location choices made over time by individuals and firms. In the US, it is well known that companies tend to locate polluting plants in the poor parts of town to minimize the political opposition from connected and wealthy households (Banzhaf et al., 2019), while house prices near polluting plants tend to be cheaper, attracting poorer individuals (Currie et al., 2015). As a result of this double selection, people near polluting plants may significantly differ from the control group along some other comorbidities risk besides pollution (Graff Zivin and Neidell, 2013).

In Denmark, these problems are less important than in the US. Incinerators tend to be located throughout the country to minimize the cost of transporting garbage to the plant. Two major plants are in Copenhagen and one, Amager Bakke, has become a ski resort in the city center. Furthermore, our rich dataset allows us to control for income, wealth, and marital status at the

individual level, with these variables coming from a government source and being measured with great accuracy. Still, we cannot rule out the possibility of systematic differences between exposed and unexposed individuals. Clearly, this possibility is highly dependent on the way the relevant samples are constructed.

Our paper is based on two main samples, related to openings (Section 5) and closings (Section 6) of incinerators respectively[21]. Our specifications rely on the comparison of exposed individuals (people who live in municipalities where incinerators were either opened or shut down) with unexposed individuals (people who live in municipalities where an incinerator has never operated, matched with the relevant exposed municipalities by population and size as measured by square kilometers; each unexposed municipality, based on the matched municipality, is assigned a fictitious opening or closing year, which is used for the sake of sample construction). To isolate the effect of incinerators on the exposed and make sure that the unexposed are never treated, we restrict the attention to the individuals who never change their parish during their lifetime. More in detail, the exposed are individuals who live in the municipalities involved in the openings or closings of incinerators, excluding the people born after the events of interest; the exposed individuals are followed up until their death, which can occur before or after the relevant opening or closing. Symmetrically, the unexposed are individuals who live in unexposed municipalities, excluding the people born after the fictitious opening or closing year; also the unexposed are followed up until death, which can occur before or after the fictitious opening or closing year[22].

Given our samples of openings and closings, our specifications might be prone to two problems. First, there can be systematic differences associated with the location of the plants (e.g., incinerators are located in areas contaminated by other sources of pollution). Second, there

[21] In this part of the paper, we cover sample construction very concisely; further details on the steps followed to construct the samples of openings and closings are specified in the sections devoted to the specifications (Section 5 and Section 6).

[22] As is explained in Section 6, while our main specification fitted to the sample of openings relies on a before-and-after comparison in a difference-in-differences framework, in the case of the closings our main regression only considers person-years observed after the closing year, actual or fictitious. The reason is that in the sample of closings our priority is to keep cumulative exposure fixed at the individual level and, accordingly, we need that the exposed individuals do not accumulate exposure to an active incinerator while being observed. Anyhow, in Subsection A.3 of the Appendix, we run a robustness check that considers also person-years observed before the closing, constructing the sample of closings in the way described above in the text, in line with the rationale we use to build the sample of openings in a difference-in-differences framework.

can be systematic differences between exposed and unexposed individuals who decide to remain in their parish during the entire sample period (e.g., people who decide to stay in a parish located in the vicinity of an incinerator might be less concerned with their health).

As regards the first issue, our specifications consider openings and closings as exogenous shocks and are designed to partial out all the unobserved characteristics related to location. Namely, in all our regressions, we control for pair of municipalities and municipality fixed effects. The first group of fixed effects ensures that comparisons are made between exposed and unexposed individuals coming from municipalities with similar population and area. The second group partials out the fixed effect of location. Evolution in exposure to background pollution common to all the citizens in the country is captured by a set of time fixed effects[23].

As for the second issue, unfortunately, in our specifications we are not allowed to control for individual fixed effects. Given that lung cancer is an event that tends to happen only once in a person's lifetime, comparing the same individual before and after the opening or the closing of an incinerator would not be informative; indeed, an individual that is observed in the post-treatment period is very likely not to have had lung cancer before; likewise, an individual who has lung cancer before the treatment is very likely not to be observed afterward. In light of this consideration, in the remainder of this subsection, we thoroughly address potential concerns related to endogeneity due to individual characteristics that are not accounted for by our specifications.

First, we rule out that exposed and unexposed individuals differ in an economically meaningful way, along our control variables. Given that people are included in our samples conditional on living in the relevant municipalities before openings and closings of incinerators, we focus on the period before the events of interest (when selection can occur) and we run T-tests comparing exposed and unexposed individuals. Table 1 shows that exposed and unexposed individuals in the sample of openings are not characterized by any economically significant differences in terms of age, income, wealth, sex, and marital status, before the opening of an incinerator. Table 2 repeats the same exercise focusing on the sample of closings. Also in this case we

[23] In the sample of openings, in an OLS framework, we control for year fixed effects; in the sample of closings, in a Cox regression framework, we consider age as the time scale and we stratify the sample by birth cohort, with combinations of age and birth cohort capturing year fixed effects.

cannot see any economically significant differences between exposed and unexposed individuals observed before the closing.

Second, we study whether the locations where openings and closings of incinerators occur are associated with different levels of lung cancer incidence prior to the events of interest, as opposed to areas unaffected by incinerators. Figure 1 is related to openings and shows the results for a linear probability model in which the dependent variable is lung cancer and the explanatory variables are dummies for pairs of municipalities as well as for the interactions between the former and a dummy that identifies the exposed individuals. The regression is run only by considering the period before the opening year (actual or fictitious); namely, the figure shows the coefficients on the interactions, which can be interpreted as the differences in lung cancer incidence between the exposed and the unexposed individuals, in each pair of municipalities, before the opening year of interest. All the pairs of municipalities but three are associated with statistically insignificant coefficients[24]. In Figure 2, we perform the same exercise by focusing on the period before the closing of an incinerator. Overall, the results are similar to those obtained in the case of the openings: in the majority of the pairs of municipalities we do not see any statistically significant differences. Both figures suggest that the two samples are balanced in terms of our dependent variable of interest as observed in the pre-treatment period[25].

Third, we analyze the presence of attrition in our samples, which is due to the fact that our identification strategy restricts the attention to the people who never change parish during their lifetime. In particular, we check that the openings and closings of incinerators do not have any significant effect on the decision to move away from areas affected by the relevant incinerators. With this aim, we construct a new version of the two samples by including all the individuals that live in the relevant municipalities (exposed or unexposed) the first time they are observed, in the period up to the relevant opening/closing year (actual or fictitious). As soon as an individual changes his/her parish, we drop all the person-years from the change on; during the year prior to the change, we turn on a dummy variable called "Mover". Figure 3 plots the average of this

[24] As for the exceptions, they are all economically small but Marjagerfjord; in this case, the opening year is 1984, meaning that the coefficient is estimated by relying on only two years (1982 and 1983), whereby we do not attribute any substantial interpretation to this outlier.

[25] Intuitively, one may expect exposed individuals observed before the closing of an incinerator to be associated with a higher risk of lung cancer as compared with the unexposed; as documented in Section 7, due to the effect of latency, this is not the case unless the intensity of exposure is particularly high.

dummy for the sample of openings, as a function of years since opening (actual or fictitious) and distinguishing between exposed and unexposed individuals; it is not possible to find any systematic and economically significant differences in the pattern followed by the average of the dummy between the two groups of individuals; importantly, the opening of the incinerator does not change this result. In Figure 4, we repeat the exercise with reference to the sample of closings, reaching the same conclusion.

Fourth, to make our analysis more granular, we investigate whether the opening or the closing of an incinerator determines the movers who live in exposed municipalities to differ from their counterparts in the unexposed municipalities. We restrict the attention to the movers identified in the previous point, as observed the year before they change their parish. To isolate the effect of the events of interest, we only consider those people who move within 5 years before and 5 years after the relevant opening or closing; indeed, it is reasonable to think that if people respond to these events by moving, they do so within a relatively short period of time. Our source of concern is the possibility that in proximity of the events that we want to investigate, exposed people with very peculiar characteristics (e.g., very wealthy) decide to leave their parish. Table 3 and Table 4 show that this is not the case by running T-tests related to our control variables, both in the 5 years before and in the 5 years after the events of interest.

We conclude that our empirical strategy is designed so that *i*) fixed effects of location are partialled out, *ii*) there appear to be no economically significant differences between exposed and unexposed individuals observed before the treatment in terms of the control variables as well as in terms of the dependent variable, *iii*) attrition has no apparent association with the events (openings and closings of incinerators) that we want to study, and *iv*) there is no economically significant difference between the exposed and unexposed individuals that we drop from our samples because they decide to leave their parish right before and right after the events that we want to study.

3.2.4 Medical Issues

Last but not least, the medical evidence suggests that there is a complex and not clearly understood interaction among cumulative exposure, intensity, latency, and cancer diagnoses.

Namely, the length of the latency period is not fixed, but can be a function of the intensity of exposure. For example, Lubin et al. (2007) study lung cancer incidence among smokers who have quit smoking. Per given amount of cigarettes smoked over time, among heavy smokers the peak of incidence of lung cancer occurs shortly after they stopped smoking, while it occurs later in life for light smokers. Since the decision to stop smoking is heavily endogenous and often triggered by a health scare, these results are not necessarily generalizable. Nevertheless, they suggest that there could be a complex interaction worth exploring.

Our strategy of analyzing closings of incinerators has indeed the aim of studying how cumulative exposure, intensity, and latency interact. During the post-closing period, the cumulative exposure of the residents is fixed and differs based on how many years an individual lived in a certain area and how exposed to pollution that area was when the plant was operating. Given this initial difference, we can follow the incidence of lung cancer diagnoses over time to identify latency. More in detail, keeping cumulative exposure fixed, the length of the latency period can be estimated as the period of time during which the exposed individuals keep experiencing an excess risk of lung cancer, despite not being exposed anymore. One can also look at how latency varies as a function of the intensity of exposure, keeping cumulative exposure fixed in this case as well.

4 Measuring Exposure to Incinerators

4.1 Air Dispersion Modelling

Our primary objective is to assess the impact on lung cancer of pollution from incinerators. Thus, we need to assign a value of exposure to every person-year included in our dataset. Given the address of person i observed in year t, annual exposure to pollution from an incinerator of interest j (x_{ijt}) depends on four variables: distance from the incinerator j (d_{ijt}), amount of pollutants emitted from the incinerator (e_{jt}), height of the chimney (h_{jt}), and weather conditions (w_{jt}). These four variables affect exposure according to an unknown functional form,

$$x_{ijt} = f(d_{ijt}, e_{jt}, h_{jt}, w_{jt}) \tag{1}$$

During our sample period, in total 72 incinerators were active in Denmark. Thus, the total exposure to incinerators associated with person-year i-t (x_{it}) depends on all the plants active in year t. As a result, all the variables mentioned above should be thought of as vectors, with each entry representing the value for a different incinerator active during the year of interest,

$$x_{it} = f(D_{it}, E_t, H_t, W_t) \tag{2}$$

Measurement of exposure poses two specific challenges. First, one needs to obtain information on the variables entering function $f(\circ)$; second, one needs to combine these variables to obtain a measure of the unobserved exposure x_{it}. In the literature on the effect of pollution on health outcomes, the most common empirical strategies adopted by economists to overcome the challenges posed by Equation 2 are two. The first involves considering the distance from the closest point source as a surrogate measure of exposure (Banzhaf and Walsh, 2008 and Currie et al., 2015). In this strategy, only one out of the four relevant variables is considered to quantify exposure and the combined effect of multiple point sources is not taken into account. The second strategy is to exploit data on the air concentration of pollutants available from public monitors (Currie et al., 2009, Moretti and Neidell, 2011, and Ebenstein et al., 2017). In this approach, the main shortcoming is that the individuals that do not live close to a public monitor are assigned a value of exposure via interpolation, which represents a rough approximation because the spatial distribution of pollution is not smooth, but characterized by jumps and discontinuities due to weather conditions. In addition, in our setting, the total air concentration of pollutants registered by public monitors would not allow us to isolate the impact exerted by emissions from incinerators. As pointed out by Hernandez-Cortes and Meng (2020), the problems characterizing the two strategies described above result in exposure mis-classification, which is a form of measurement error and can lead to biased results.

An alternative strategy addressing some of the challenges mentioned above is to use of an air dispersion model. In a nutshell, an air dispersion model takes as inputs the coordinates of multiple point sources, their emissions, their characteristics (mainly the height of the chimney), and weather conditions (mainly wind direction); the output is the spatial distribution of the concentration of the pollutants of interest. Since the 2000s, air dispersion models have become

the standard approach to measure air pollution consequences in the environmental and medical science literature (Næss et al., 2007, Pennington et al., 2018, and Filigrana et al., 2022). In contrast, only few economists have taken advantage of this tool to measure exposure. The most important references in economics are Sullivan (2017), Grainger and Ruangmas (2018), Hernandez-Cortes and Meng (2020), and Christensen et al. (2022). These studies belong to the research strand devoted to environmental justice, aiming to assess whether specific geographical areas (e.g., ZIP codes) populated by racial minorities and low-income households are particularly exposed to toxic pollution. To the best of our knowledge, our paper is the first in economics to combine the use of an air dispersion model with individual-level data on residence as well as on economic and health status to study long-term health effects of exposure to pollutants. Barreca et al. (2021) work in a similar setting, but their analysis relies on aggregate economic and health data at the county level and not on the exposure history reconstructed at the individual level.

4.2 HYSPLIT

The two models of air dispersion most commonly used in the literature are AERMOD[26] and HYSPLIT[27], developed respectively by the US EPA and by the National Oceanographic and Atmospheric Administration. Between AERMOD and HYSPLIT, we decide to use HYSPLIT to model the air dispersion of annual emissions from incinerators. AERMOD provides a more rigorous modelling of the dispersion of pollutants in the atmosphere, taking into account chemical reactions between different pollutants and the relationship between each substance and its precursors; nevertheless, this model is not computationally agile and is typically used by regulatory agencies to perform short-range analyses (at most 50Km) based on one or a few pollution sources. In contrast, HYSPLIT, thanks to its ability to balance computational tractability and scientific realism, is more suitable for our large-scale analysis that involves several point sources every year (Hernandez-Cortes and Meng, 2020).

We run HYSPLIT by adapting the procedure proposed in Henneman et al. (2019) to our framework[28]. This procedure involves simulating the release of 100 fictitious particles of a

[26] https://www.epa.gov/scram/aermod-modeling-system-development
[27] https://www.ready.noaa.gov/HYSPLIT.php
[28] To run HYSPLIT, we make use of the R package called DisperseR. In addition, we exploit the resources

pollutant of interest (in our case, primarily SO2) from each incinerator-year between 1975 and 2018, 4 times a day (12am, 6am, 12pm, and 6pm). Each release is tracked for a number of hours that depends on the pollutant under investigation (according to Hernandez-Cortes and Meng, 2020, 13 hours are needed for SO2), taking into consideration the height of the chimney of the relevant incinerator as well as hourly weather conditions provided by the National Oceanographic and Atmospheric Administration reanalysis database. The results obtained from daily simulations yield the annual concentration of the particles emitted from each incinerator with reference to some geographical units, parishes in our case. Then, for each parish, the weighted sum of the concentrations associated with different incinerators is computed, considering annual emissions of the pollutant of interest as weights. The resulting unitless measure is called HyADS (short for HYSPLIT Average Dispersion) and ranks parish-years based on exposure to incinerators. Depending on the parish where person i lives in year t, we compute x_{it} as the value of HyADS associated with the relevant parish-year. This allows us to overcome the difficulties arising from Equation 2.

4.3 SO2 as a Proxy for Pollution from Incinerators

A key empirical choice in our analysis is to select the pollutants of interest (or the best proxies for such pollutants) and to measure annual emissions of such pollutants from incinerators, which represent the main input used to run the dispersion model. Since our main objective is to measure the long-term effect of exposure to incinerators on lung cancer, we need to proxy for PM2.5, dioxins, and heavy metals, which are the pollutants emitted from incinerators that have been linked to lung cancer by the International Agency for Research on Cancer (IARC)[29]. Among these substances, we are especially interested in finding a good proxy for PM2.5, because of two reasons. First, stationary combustion plants (including incinerators) represent the largest source of PM2.5 (Krause and Smith, 2006) and this substance cannot be captured by state-of-the-art filters (Buekens, 2013), representing a problem even for modern plants. Second, PM2.5 has been recognized as a human carcinogen with the safest level of scientific evidence (Group 1) by

provided by the Mercury high-performance computing cluster at The University of Chicago Booth School of Business, which is supported by the Office of the Dean.

[29] https://monographs.iarc.who.int/agents-classified-by-the-iarc/

the IARC.

In our main analysis, we decide to use SO2 as a proxy for the entire toxic pollution mix (PM2.5, dioxins, and heavy metals) emitted by incinerators. We justify this choice by resorting to two reasons. First, SO2 is the substance for which we have the most complete data on actual emissions from the available Green Accounts. Second, SO2 is strongly correlated with other air pollutants, performing very well as a general pollution tracker and being used since the 1970s as a criterion pollutant by regulatory agencies all over the world, including the US EPA. In particular, SO2 is closely linked to PM2.5, being one of its precursors in the atmosphere (Henneman et al., 2019 and Barreca et al., 2021)[30].

4.4 Assigning SO2 Emissions to Incinerator-Years without Green Accounts

While the Green Accounts are the best source of data on annual SO2 emissions, they are available from 1996 (and at the moment we have access only to reports published in the periods 1996-2001 and 2007-2018), with some incinerator-years missing due to the confidentiality of their reports or to size thresholds established by law. In total, out of the 1,526 incinerator-years included in our dataset spanning 1975-2018, we observe SO2 emissions from the Green Accounts for only 451 of them. To supplement the data on SO2 for the incinerator-years without Green Accounts, we use two additional data sources: *i*) imputed emissions computed as the product between the annual sector-specific SO2 emissions factor and the annual amount of waste burnt; *ii*) annual concentration of SO2 estimated by the NASA in correspondence to the parish where each incinerator of interest is located. To study the relationship between actual SO2 emissions and these other variables, in Table 5, we regress actual emissions from the Green Accounts on imputed emissions and NASA concentrations of SO2 at the parish level. In Table 5, (III) both variables enter the regression with statistically significant coefficients associated with

[30]Consistent with the literature, we find that the correlation between SO2 and PM2.5 concentrations measured by the monitors of the University of Aarhus is 0.804 and the correlation between the same concentrations estimated by the NASA is 0.850, with the p-value being lower than 1% in both cases. In future analysis, we also plan to compute versions of HyADS based on estimated emissions of PM2.5, dioxins, and heavy metals. These versions rely on estimates of these pollutants, as they are not present in the Green Accounts. The details are in Subsection A.2 of the Appendix.

the expected sign; the R^2 is 0.42.

The results presented in Table 5 show that both imputed emissions and NASA concentrations perform well as predictors of actual emissions of SO2. We use them both to assign emissions of SO2 to the incinerator-years without Green Accounts, exploiting the fitted values from Table 5, (III)[31]. Figure 5 shows the annual sum of SO2 emissions (in tonnes) between 1975 and 2018, distinguishing between actual and fitted emissions[32].

4.5 Computing HyADS Based on SO2 Emissions

Using fitted and actual emissions of SO2, we run HYSPLIT and compute a value of HyADS for each parish-year between 1975 and 2018. Figure 6 shows the map of Denmark in 1975, in 1997, and in 2018. HyADS is characterized by a strong variation, both in the time series (the peak is 6.340 in 1975 and 0.266 in 2018) and in the cross-section. Over time, new sources of pollution arise; for example, in 1997, the Naestved plant turns out to be the most polluting incinerator, emitting 380 tonnes of SO2 according to its Green Account.

To show that our measure is a good proxy for exposure to pollution from incinerators, we regress the concentrations of SO2 registered by the monitors of the University of Aarhus on HyADS based on actual and fitted emissions of SO2 and computed with reference to the parishes associated with the monitors. According to Table 6, HyADS alone explains more than half of the variance of the dependent variable. The coefficient on HyADS remains positive and statistically significant after controlling for year and monitor fixed effects. Considering that HyADS is only based on pollution from one source (incinerators) and that monitors are not necessarily located in the vicinity of the plants of interest, these results suggest that HyADS is a good proxy for exposure to incinerators over time and across different areas.

[31]While imputed emissions are designed to proxy for actual emissions and NASA concentrations are not specifically linked to incinerators as a source of pollution, we use both variables due to the availability over time of the quantities used to compute imputed emissions. More details on the justification of this choice are provided in Subsection A.1 of the Appendix.

[32]In order to compute fitted values, we need to make some assumptions on imputed emissions and NASA concentrations due to data availability. Namely, for the period prior to 1990, we assume that the 1990 emissions factors stayed constant; for the period prior to 1994, we discount the 1994 data on waste burnt by real GDP growth; for the period prior to 1980, we assume that the 1980 concentrations stayed constant.

4.6 Cumulative Exposure to Pollution

Given that we are interested in the effects of long-term exposure to pollution from incinerators, we need to compute a measure of lifetime cumulative exposure with reference to each person-year included in our dataset spanning the period 1982-2018, when all our controls are available. We face the problem of attributing exposure prior to 1975. Indeed, individual-level data on residential addresses are available for the period between 1975 and 2018, which is the time interval for which we run HYSPLIT. In addition to this information, we only know when the people observed in 1975 moved to their address for the first time.

As a first step, we compute the number of HyADS units for each parish observed before 1975 by assuming that the number of HyADS units generated by each plant active in 1975 remained constant for the period before 1975[33]; as we go back in time, less incinerators contribute to the number of HyADS assigned to each parish, up to 1950, when the first incinerator included in our dataset was activated. Thanks to this assumption, we are able to associate each parish-year between 1950 and 2018 with a number of HyADS units. Given this, we adopt the following strategy to compute cumulative exposure with reference to an arbitrary person-year i-t observed between 1982 and 2018:

- From 1975 to t, we consider the cumulative sum of HyADS computed at the parish level, based on the residential address of i observed every year.
- Between 1975 and the moving year associated with the 1975 address, we consider the cumulative sum of HyADS observed with reference to the parish of the 1975 address.
- Between the moving year and the most recent between the birth year and 1950, we discount the annual values of HyADS by considering 1975 as the reference year and using a discount rate of 0.9, justified by an annual moving rate of 10%[34].

[33] We make this choice because for the period before 1975 we have very poor information on the predictors used to assign emissions to incinerator-years without Green Accounts, imputed emissions (emissions factors are available from 1990 and the amount of waste burnt is observed from 1994) and NASA concentrations of SO2 (available from 1980). While for the period between 1975 and 1994 we make the assumptions explained above in the paper, for the period before 1975 (when we do not even have residential addresses), we decided not to run the dispersion model based on highly uncertain information on emissions.

[34] 10% is obtained as the sample average of a dummy variable that takes the value of 1 with reference to person-year i-t if i is associated with two different parishes in $t-1$ and t.

To provide an example, with reference to a person i observed in 1985, who was born before 1950 and moved to her 1975 address in 1970, cumulative HyADS is computed as

$$\text{Cum.HyADS}_{i,1985} = \sum_{t=1975}^{1985} \text{HyADS}_{i,t} + \sum_{t=1970}^{1974} \text{HyADS}_{i,t} + \sum_{t=1950}^{1969} 0.9^{1974-t}\text{HyADS}_{i,t} \quad (3)$$

where the last two summations are related to the parish associated with the address of i in 1975. Applying Equation 3 to our master dataset, we obtain the distribution for cumulative HyADS presented in Table 7.

5 Openings

Based on the medical literature on cancer development (e.g., Pitot, 1993), we know that the relationship between exposure to incinerators and lung cancer, if any, should not be contemporaneous, but characterized by a delay. Thus, our first empirical challenge is to determine the presence of a relationship between exposure and lung cancer, measuring the delay between the start of exposure and the manifestation of the disease.

To assess the presence of a causal relationship and of a delay, ideally, one would like to follow two cohorts of individuals equal in all respects except for the fact that they live in two different places far apart; in turn, the two places should be perfectly equal, except for the fact that one of them starts being exposed to emissions from an incinerator at a relatively early stage of the life-cycle of the two cohorts. Since then, one should record the health status of the two cohorts year by year and compare the frequency of cancer diagnosis. Assuming that the two cohorts do not mix with each other (i.e., they do not move), the time profile of the difference in cancer diagnoses between the two cohorts in the years after treatment identifies the causal effect of emissions and pins down the delay between the start of exposure and cancer development. Furthermore, if members of the treated cohort are differentially exposed to emissions (e.g. because some of them live closer to the source), one can trace the interaction between intensity at the start of exposure and delayed response.

The openings of incinerators can be thought of as exogenous shocks to the exposure of people living in a vicinity of the new plants, representing the start of exposure for these people. Below,

we describe how we exploit openings to mimic our ideal set up in a difference-in-differences framework.

5.1 Design of the Event Study

The first step to construct our sample is to consider only the individuals that never change parish (the geographical unit in correspondence to which exposure to incinerators is measured) during their entire lifetime.

Given the first restriction, we define exposed individuals as those who live in a "virgin municipality" (one with no incinerators) affected by the opening of an incinerator during the period for which we have the entire set of control variables (1982-2018). In total, during this period there are 15 openings, the first occurring in 1982 the last in 2003 (see Table 8), meaning that for the oldest opening we can follow people for up to 36 years and for the most recent one for up to 15 years. Our exposed individuals must live in the relevant municipalities in the years up to the opening and are followed up until death/end of the sample period in 2018. In this manner, individuals who die or incur in a cancer diagnosis before the opening year are included among the exposed individuals and contribute to the before-and-after comparison. Instead, individuals born after the opening year are not considered.

Each exposed municipality is matched with a municipality associated with no incinerators, choosing the closest municipality based on population and size (measured by the area in square kilometers of the municipality). Namely, referring to an exposed municipality i, we select the matched unexposed municipality j from the set J of municipalities associated with no incinerators by applying the following formula,

$$j = \min_{J} \left\{ \frac{|\text{Population}_j - \text{Population}_i|}{\text{Population}_i} + \frac{|\text{Area}_j - \text{Area}_i|}{\text{Area}_i} \right\} \tag{4}$$

Based on the corresponding exposed municipality, each one of the unexposed municipalities is associated with a fictitious opening year - the same year as the opening year in the treated group. The fictitious opening year is then used to define the control group; this is made of the collection of individuals who have been living in an unexposed municipality during the period up to the

fictitious opening year, including those who die or incur in a cancer diagnosis before. We drop all the newborns after the fictitious opening year. Hence, the control group follows exactly the same selection rules as the treated group.

Clearly, even though people in the control group live in a municipality with no incinerators, they may suffer some exposure because emissions travel with the air. To come as close as possible to an ideal control group, we drop from the control group individuals who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. For symmetry, from the treated group we drop individuals who are associated at least once with a level of HyADS of at least 0.9 units before the opening year of the incinerator of interest. The use of a threshold in terms of HyADS units is needed to make sure that the untreated are constantly associated with a relatively low level of exposure and that the treated are not associated with a significant level of exposure before the opening, which should represent a clear shock to their exposure history. We choose 0.9 units as the cutoff because it identifies the top 30% of the exposed individuals according to the distribution of HyADS at the time of the opening of an incinerator. As shown in Figure 7, this group of exposed individuals experiences the most significant shock at the time of the opening. In addition, thanks to the 0.9 threshold, we obtain a sample that is relatively balanced in terms of number of treated and untreated individuals: in total, we end up with 6,144,782 person-years, with the exposed accounting for 3,497,842 observations.

From Figure 7, we can infer that the event study is well-structured. Not only there are not economically significant differences between the treated sample and the control group (as shown by Table 1), also HyADS is very stable both for exposed and unexposed individuals before the opening year. In correspondence to the opening year, the exposed individuals (especially the top deciles) experience a neat jump in their level of exposure, while the unexposed do not experience any change. It is also interesting to notice that the size of the shock hitting the exposed individuals depends on the technology vintage of the incinerators to which they are exposed. In Table 9, we show the median opening year for the incinerators associated with the deciles of exposed individuals ranked by HyADS at the time of the opening; while the first decile has 1995 as the median opening year, all the deciles in the top 50% of the distribution are associated with 1984. This pattern clearly reflects the fact that incinerators have significantly

reduced their emissions over time, as an effect of new technology and emissions caps introduced by environmental regulation (see Subsection 2.2).

5.2 Results

In our specification, we follow a difference-in-differences framework, where lung cancer incidence is the dependent variable in a linear probability model. The main explanatory variables are categories of years since opening, turned on only with reference to the exposed individuals. We control for municipality fixed effects (which also absorb the fixed effect of being exposed), pair of municipalities fixed effects (to make sure that comparisons are made with reference to exposed and unexposed individuals coming from the same pair of matched municipalities), and year fixed effects. In addition, we control for lagged income and wealth quintiles, sex, marital status, and dummies for age ranges. We run the same specification four times. First, we estimate the model for the entire sample; second, we select only the bottom three deciles among exposed individuals (HyADS$<$0.4 at the time of the opening); third, we consider the sample containing only the middle 4 deciles (HyADS between 0.4 and 0.9 at the time of the opening); fourth, we restrict the attention to the top three deciles (HyADS$\geq$0.9 at the time of the opening). In the latter three cases, the control sample of unexposed individuals is obtained by applying the pairwise matching at the municipality level only with reference to the exposed individuals included in the deciles of interest.

In Table 10, (I), considering the entire sample, the main result that we obtain is that, after the start of exposure, the excess risk of lung cancer for the exposed materializes with a decade-long delay (25-30 years); the magnitude of the excess risk is significant, both statistically and economically (4 basis points, with the sample average of lung cancer being 23 and the average related to the person-years observed 25 years after the opening, actual or fictitious, being 35).

In Table 10, (II), we show that for the individuals with the lowest level of exposure at the time of the opening, lung cancer incidence increases only after at least 30 years since the opening[35].

[35]Note that we have a data problem for the coefficient associated with the 30+ Years Since Opening category, because that coefficient is estimated with reference to only one incinerator (Nykobing F), which is the only plant opened early in the sample and associated with low emissions; removing this plant from the sample of the bottom three deciles makes it impossible to estimate the coefficient for 30+ Years Since Opening and does not substantially change the other parameter estimates.

In Table 10, (IV) we document that for the individuals with the highest level of exposure at the time of the opening, the excess risk of lung cancer manifests itself right after the opening, even though it reaches its peak after around 10 years. In Table 10, (III), focusing on the middle 40% of exposed individuals, we find intermediate results between the two extremes represented by (II) and (IV).

The implication of Table 10, (I) is that there exists a relationship between exposure to incinerators and lung cancer and is characterized by the presence of a decade-long delay between the start of exposure and the development of lung cancer; the presence of such delay might be due to the role played by cumulative exposure (i.e., individuals need to accumulate significant amounts of exposure before developing the disease) or latency (i.e., the effect of exposure on lung cancer materializes after a latency period). The implication of Table 10, (II), (III), and (IV), is that the delay is much shorter for the group of individuals with the highest exposure at the time of the opening. Also this pattern could be the result of two effects. On the one hand, individuals included in the top deciles accumulate exposure at a higher rate, whereby their level of cumulative exposure grows faster and they are soon exposed to large dosages; on the other hand, it is possible that high-intensity exposure reduces the latency period, making lung cancer incidence increase earlier. To shed light on these effects, we move on to the following section, devoted to the closings.

6 Closings

In the sample of openings, individuals accumulate exposure to pollution while being observed and this makes it impossible to distinguish between cumulative exposure and latency. In addition, intensity has been assessed only with reference to the time of the opening, without considering the entire exposure history of the individuals of interest. To distinguish between cumulative exposure and latency, we need a setting in which we can compare unexposed and exposed individuals, where the source of pollution is exogenously removed, and both treated and control groups are followed after the removal. We use the closing of incinerators as the exogenous event to study how lung cancer evolves as a function of time since closing, keeping cumulative

exposure constant. In this setting, exposed individuals are observed after they stop accumulating exposure to incinerators, due to the closing of the incinerator located in their municipality. Besides latency and cumulative exposure, in our setting, we can also model a third dimension, the intensity of exposure, which should play an important role and interact with the other two dimensions according to the medical literature (Lubin et al., 2007).

6.1 Construction of the Sample

The construction of the sample of closings follows a rationale similar to the one applied to the openings.

Also in this case, the first step involves restricting the attention to the people who never change parish during their entire lifetime. We consider the municipalities where an incinerator was shut down between 1982 and 2018 and where, after the closing of the incinerator of interest, no new plants have ever been built. Overall, there are 20 of such municipalities, with the first closing occurring in 1985 and the last one in 2017 (see Table 8). Exposed individuals are identified as people who live in one of these municipalities, during the relevant closing year.

We match each of the 20 exposed municipalities with one municipality where an incinerator has never operated, considering population and area according to the same formula used for the openings (Equation 4). Each unexposed municipality is associated with a fictitious closing year, based on the matched exposed municipality. Unexposed individuals are people who live in unexposed municipalities, during the relevant fictitious closing year.

We drop unexposed individuals who are associated at least once with a level of HyADS of at least 0.9 during their entire lifetime and exposed individuals who are associated at least once with a level of HyADS of at least 0.9 after the relevant closing year. Finally, we start observing exposed and unexposed individuals after the closing year (actual or fictitious) and we end up with a sample consisting of 3,015,359 person-years, with the exposed accounting for 1,906,997 observations[36].

The main difference from the sample of openings is that in this case we observe the individuals of interest only after the event under investigation. The reason is that we want to make sure

[36] In Table 11 and Table 12 the number of observations is only 3,009,360 because in a Cox regression one has to drop the post-failure observations (i.e., the person-years after the first diagnosis of lung cancer).

that cumulative exposure to incinerators is fixed for every individual, so that we can separate its effect from the effect of latency[37].

6.2 Model Specification

The specification of our empirical model is an adaptation of that estimated in Lubin et al. (2007) to our framework. Namely, Lubin et al. (2007) are interested in modelling the interactions among cumulative exposure to tobacco smoke (measured in pack-years, with 1 unit representing the amount of cumulative exposure equivalent to smoking one pack of cigarettes every day for one year), intensity of exposure (average number of cigarettes smoked per day), and latency (time since cessation) as predictors of lung cancer for former smokers. In their specifications, cumulative exposure is treated as a continuous variable and the effect of one more pack-year is let vary with combinations of categories of years since cessation and categories of numbers of cigarettes smoked per day. We map this empirical framework into ours. For each exposed individual, we compute the lifetime cumulative sum of HyADS units up to the time of closing. We use this quantity as our measure of cumulative exposure, we divide it by 0.9 (the threshold identified before) and consider 0.9-HyADS-Years as our measurement unit. Our corresponding measure of latency is years since closing. Finally, we consider the cumulative sum of HyADS at the time of closing divided by age at closing as our measure of intensity of exposure[38].

Given this setting, we run two Cox regressions. In the first, we interact cumulative exposure with four categories of years since closing (0-4, 5-9, 10-14 and 15+ years). In this specification, we capture the interaction between latency and cumulative exposure. We expect exposed individuals to experience an excess risk of getting cancer even though they are observed *after* they stop accumulating exposure - a latency effect. In addition, we expect the impact of one more unit of cumulative exposure on lung cancer to be positive - an effect of dosage.

In the second Cox regression, we let the effect of 1 more 0.9-HyADS-Year vary not only with the number of years since closing, but also with average intensity, captured with three categories that match the categorization performed in the sample of openings (<0.4, 0.4-0.9, and

[37] In Subsection A.3, this restriction is relaxed.

[38] Note that our data on cumulative exposure to incinerators span the period 1950-2018, as explained in Subsection 4.6. Thus, if an individual was born before 1950, to compute average intensity we use the difference between the closing year and 1950 as the denominator.

$\geq$0.9 HyADS units). This specification allows us to assess whether the rate at which individuals accumulate exposure has an effect on the overall increase in lung cancer risk and on the length of the latency period.

In both Cox regressions, we control for income and wealth quintilies lagged by two years as well as for marital status, we consider age as the time scale, and we stratify the sample by sex, birth cohort, and pair of municipalities.

6.3 Results

Table 11 shows the results for the first Cox regression, in which we only consider cumulative exposure and time since closing. Each hazard ratio must be interpreted as the gross excess risk associated with 1 more 0.9-HyADS-Year of cumulative exposure before the closing, accounting for the number of years elapsed since the closing. The results imply that the effect of cumulative exposure to an incinerator persists for 15 years after its closing, pointing at a 15-year latency period. Considering that the hazard ratios related to the first 15 years after the closing are equal to around 1.005 and that the average exposed individual has an amount of cumulative exposure at the time of the closing of 15.51 0.9-HyADS-Years, we can conclude that, on average, the exposed are more likely to get lung cancer as opposed to the unexposed according to an overall hazard ratio of $1.005^{15.51} = 1.08$ in the relevant time interval.

In Table 12 we run the previous regression allowing the estimates of the latency period to differ according to average exposure intensity. The results indicate that for low levels of intensity (0-0.4 HyADS) the latency period is of 10-15 years; for intermediate toxic exposures (0.4-0.9 HyADS) the latency period falls to 5-15 years; finally, for the highest category of intensity (≥ 0.9 HyADS), the bulk of the effect is concentrated within 5 years since the closing. In sum, our evidence implies that the length of the latency period is inversely related to the intensity of exposure.

In Section 7, based on the hazard ratios estimated in Table 12, we perform back-of-the-envelope calculations that clarify how our hazard ratios must be interpreted, quantify the impact of exposure to incinerators on lung cancer, and assess the importance of accounting for latency. For robustness, in Subsection A.3, we fit a Cox regression to the sample of closings in which

we observe both exposed and unexposed individuals also before the closing year (actual or fictitious), including in the sample also people who die or incur in lung cancer before this event. Table A.4 shows that our results are robust; for the post-closing period, we find the same pattern as in Table 12, as expected; before the closing, we find that only the individuals associated with the highest average intensity incur in an excess risk of lung cancer.

6.4 Unobserved Heterogeneity

In all our estimates, we control for financial and demographic characteristics, which are very good predictors of risky behavior associated with negative wealth outcomes. However, since we do not observe directly smoking behavior, the most powerful determinant of lung cancer, we worry that our estimates are potentially prone to a problem of unobserved heterogeneity. If smoking habits were to be correlated with exposure to emissions from incinerators, our estimates may be picking the former rather than the latter. In this section, we address this issue in two robustness exercises where we verify *i*) whether smoking habits are positively correlated with exposure to incinerators and *ii*) whether the effect of exposure to incinerators on lung cancer incidence vanishes after controlling for smoking habits.

To implement these strategies, we rely on a the sample of all pregnant women observed between 1999 and 2018 (1,199,719 observations). Each pregnant-woman-year is associated with a dummy ("Smoker") that takes the value of 1 if the woman is classified as a current smoker. In Table 13, we regress Smoker on HyADS measured at the parish-year level. In Table 13, (I), controlling only for time fixed effects to account for common trends and municipality fixed effects (which are the main fixed effects included in our regressions fitted to openings and closings), HyADS and Smoker are negatively correlated. In Table 13, (II), adding the full set of individual controls - log income and log wealth, age and dummies for marital status - amplifies the negative correlation. Hence, there seems little support for the idea that the effects of emissions on lung cancer may be due to omitted smoking habits. Importantly, when we add all the individual controls, the fit of the regression increases substantially (the R^2 goes up from 0.029 to 0.081), implying that even if we do not control directly for smoking habits, variation in the latter is well captured by our controls, particularly by income and wealth.

Having verified that the correlation between HyADS and Smoker is, if anything, negative, we check whether accounting for smoking can dampen the effect of exposure to incinerators on lung cancer. Given that a woman is very unlikely to have lung cancer while pregnant, to gain statistical power, we do not limit our analysis to pregnant-woman-years, but consider all the person-years from our master dataset associated with the women who have been pregnant between 1999 and 2018 (18,854,406 observations). Then, we define a dummy called "Ever Smoker" that takes the value of 1 if a woman is ever classified as a smoker during the sample period and is fixed for the same woman. Our independent variables of interest are Ever Smoker as well as the quartiles of cumulative HyADS (the first quartile is omitted and considered as the baseline), measured according to the methodology explained in Subsection 4.6; the other controls are age range, lagged income, and wealth quintiles, marital status, year and municipality fixed effects. Estimates are shown in Table 14. In Table 14, (I) we include all controls except Ever Smoker; there is a monotonic relation between cumulative HyADS and lung cancer, with the top quartile being associated with a coefficient of roughly 0.3 basis points, the same size as the baseline probability of lung cancer in this sample. In Table 14, (II) we add the Ever Smoker control. Smoking habits have a positive and highly significant effects on the probability of lung cancer, both statistically and economically (the coefficient is 0.5 basis points). However, their inclusion does not substantially modify the OLS coefficients of the cumulative HyADS; namely, no one of the coefficients on cumulative HyADS is statistically different from Table 14, (I); the only coefficient that loses statistical significance is the one on the second quartile of HyADS; however, even in Table 14, (I), this coefficient is significant only at the 10% level, with the p-value being 0.098. Moreover, including Ever Smoker changes the R^2 of the estimated model only by roughly 0.1 basis points, from 1.4 to 1.5[39].

[39]Note that in Table 14 we do not account for latency and cumulative exposure enters the regression in a contemporaneous fashion. The reason is that in this framework we are not interested in capturing the true impact of exposure to incinerators on lung cancer, but we only want to check whether controlling for smoking habits kills the association between HyADS and the dependent variable.

7 Discussion

Our results from Table 12 uncover the importance of the interaction among cumulative exposure, intensity, and latency in the relation between toxic pollution and lung cancer. In this section, based on our results, we try to assess the underestimation of the health effects of pollution due to empirical strategies that do not take latency into account. We do so both by focusing on the individuals and on the incinerators included in the sample of closings (from which the estimates are drawn) and by running a simulation based on an incinerator that pollutes at an intensity equal to the one currently associated with the plant located in Copenhagen. Finally, thanks to our estimates, we are also allowed to quantify the impact of regulatory changes on the health effects of incinerators over time.

7.1 Assessing the Underestimation of the Health Effects of Pollution

7.1.1 Sample of Closings

We first focus on the sample of closings. The estimates from Table 12 allow us to quantify the excess risk of lung cancer experienced by people who have been exposed to an incinerator and are observed during the post-closing period. Thanks to our specification, we can distinguish among different categories of intensity of exposure and different categories of years since closing, while keeping cumulative exposure fixed and adjusting for our covariates.

In the Appendix, Subsection A.4, we estimate that during the post-closing period 410 excess lung cancer cases can be observed among the exposed individuals (149,539 unique people and 1,907,181 person-years) as a result of cumulative exposure accumulated before the closing. The temporal distribution of these excess cases of lung cancer depends on the intensity of exposure, as the hazard ratios in Table 12 document. As a result, only 82 (all related to the highest category of intensity) out of the 410 lung cancer cases manifest within 5 years since the closing. Assuming that these 82 cases are uniformly distributed over the first 5 years since the closing, we estimate that ignoring the role played by latency in the identification strategy (i.e., limiting the analysis to the closing year, under the assumption that the effect of cumulative exposure vanishes after people stop being exposed) would lead to estimate only 4% (16 out of 410) of the

overall number of excess lung cancer cases caused by exposure accumulated before the closing and occurred in the post-closing period. The underestimation is 100% for the first two categories of intensity (up to 0.9 HyADS) and equal to 90% for a level of intensity above 0.9 HyADS.

7.1.2 Simulation

In addition to performing back-of-the-envelope calculations with reference to the sample of closings, we run a simulation to have a better idea of the empirical challenges posed by the phenomena uncovered by our study. We consider two thought experiments, referring to an econometrician interested in the lung cancer implications of exposure to pollution from an incinerator of interest. First, the econometrician wants to estimate the impact of the incinerator on lung cancer, observing individuals exposed to the plant for 0, 10, 20, and 30 years starting from the opening. Second, an environmental policy implemented 30 years after the opening of the plant cuts its emissions to zero and the econometrician wants to assess whether the policy has had any beneficial impact on lung cancer incidence for the exposed individuals.

In order for our thought experiments to be especially relevant to the current debate around waste incineration, we consider an incinerator that pollutes at an intensity belonging to the first category (0-0.4 HyADS), which is the median category of the intensity of exposure for exposed individuals included in the sample of closings; in particular, we choose 0.2 HyADS, which happens to be the number of HyADS units currently generated every year by Amager Bakke, the Copenhagen incinerator, with reference to its parish. This incinerator was opened in 2016 and is praised as the most modern and environmentally friendly facility in the world. For the remainder of our simulation, we will use the result that, for the relevant category of intensity, we have estimated a hazard ratio of 1.0144 for every additional 0.9-HyADS-Year of cumulative exposure, with the latency period being 10-15 years (Table 12).

We make some assumptions needed to simplify our simulation. Three assumptions are related to the modelling of latency. *i*) We assume the latency period to be equal to 10 years, which is the lower bound of the range we have estimated. *ii*) We define latency such that the effect of cumulative exposure on lung cancer materializes with a delay exactly equal to the latency period; this means that the risk of lung cancer for an individual i observed at time t is

affected by the amount of cumulative exposure that i had at time $t-10$ and that cumulative exposure between $t-9$ and t has no impact on the risk of lung cancer at time t. *iii*) The effect of cumulative exposure to pollution completely vanishes after a period of 10 years of absence of exposure. The latter assumption is in line with the literature on the effect of cancer promoters (e.g., Weisburger, 1990); according to this literature, indeed, after a long period without exposure, the excess risk of cancer generated by a cancer promoter should vanish. In addition, our assumption is in line with our estimates, because we find that the effect of cumulative exposure disappears some years after the closing. We also assume that the exposed individuals observed by the econometrician are all associated with 0.2 HyADS units per year during the first 30 years of activity of the incinerator and with 0 HyADS after the implementation of the policy; the unexposed individuals who form the control group used by the econometrician are constantly associated with 0 units of HyADS, both before and after the implementation of the policy. Every individual is assumed to never change her residential address. Finally, we assume the baseline risk of lung cancer to be constantly equal to 20 basis points, based on the sample average (22.58) computed with reference to the sample of closings.

As for the first thought experiment, in years 0-9 since the opening of the plant, the econometrician will be able to detect no effect on lung cancer for exposed individuals, who are associated with no 0.9-HyADS-Years of cumulative exposure 10 years before. In year 10, the exposed individuals have accumulated $0.2 \cdot 10/0.9 = 2.22$ 0.9-HyADS-Years of cumulative exposure; nevertheless, due to the presence of latency, the observed annual excess lung cancer risk only depends on exposure accumulated during the opening year (10 years before) and is equal to $20 \cdot (1.0144^{(0.2/0.9)}-1) = 0.06$ basis points: unless the sample is extremely large, the econometrician would not be able to find a statistically significant estimate of this excess risk, which means that in year 10 not accounting for latency would lead to a 100% underestimate of the consequences of cumulative exposure to the incinerator. Let us move to year 20. At this point, the econometrician should observe $20 \cdot (1.0144^{(0.2 \cdot 10/0.9)}-1) = 0.65$ basis points of excess risk for the exposed individuals; nevertheless, in a sample of, let's say, 10,000 exposed individuals, the econometrician is likely to capture nothing at all; in addition, even if she captured the effect, this would only be equal to less than 50% of the true effect of cumulative exposure up to that time,

$20 \cdot (1.0144^{(0.2 \cdot 20/0.9)} - 1) = 1.31$ basis points of excess lung cancer risk. Let us move to year 30. At this point, due to latency, the econometrician observes $20 \cdot (1.0144^{(0.2 \cdot 20/0.9)} - 1) = 1.31$ basis points of excess risk; while an excess risk of 1.31 basis points is quite likely to be detected empirically, it still represents roughly 66% of the true effect of cumulative exposure at that time, which is of $20 \cdot (1.0144^{(0.2 \cdot 30/0.9)} - 1) = 2.00$ basis points. We conclude that according to our estimates from the sample of closings, to find something, the econometrician needs to observe the exposed individuals in a panel dataset that covers around 30 years starting from the opening of the incinerator. This conclusion is also in line with the estimates found in the sample of openings (Table 10, I), even though we ran the simulation according to the hazard ratios found in the sample of closings.

As for the second thought experiment, after 30 years of activity at 0.2 HyADS per year, the incinerator under investigation reduces its emissions to zero thanks to the implementation of the environmental policy under investigation. The econometrician wants to assess the effect of this policy by means of a difference-in-differences estimator. Before the reduction in emissions, the exposed individuals had 1.31 basis points of excess lung cancer risk as opposed to the unexposed individuals, as explained in the first thought experiment. 10 years after the policy, the econometrician would observe an excess risk of $20 \cdot (1.0144^{(0.2 \cdot 30/0.9)} - 1) = 2.00$ basis points, because she will be looking at the effect of 30 years of cumulative exposure to the plant, as a consequence of the 10-year latency period. Thus, a difference-in-differences estimator relying on only 10 years of follow-up during the post-treatment period may lead the econometrician to conclude that the policy has had no effect on lung cancer incidence; if any, the effect estimated by the difference-in-differences estimator would point at the fact that the policy has increased excess lung cancer incidence by $2.00 - 1.31 = 0.69$ basis points. For the estimator to document the true effect of the policy, the follow-up after the policy must be very long, more than 10 years; indeed, in our setting, it is after more than 10 years that excess lung cancer cases will vanish as an effect of the policy.

We infer that according to both thought experiments empirical strategies must be driven by the awareness of the role played by latency and supported by very long panel datasets. Otherwise, estimates of the damages of pollution as well as of the benefits of environmental policy can be

remarkably misleading.

7.2 Assessing the Benefits of Regulation

Finally, exploiting the estimates drawn from the sample of closings, we attempt to assess incinerators as a source of pollution in Denmark and to describe the evolution in their health effects over time, after regulation curbing emission was passed. As documented in Subsection 2.2, incinerators were subjected to two emissions caps, in 1991 and 2003. In light of this consideration, it is reasonable to expect the health consequences of the activity carried out by incinerators to have declined over the years. Based on the hazard ratios shown in Table 12, we consider our entire dataset covering the universe of Danish people and incinerators observed between 1982 and 2018 with the aim of assessing how many excess lung cancer cases the median incinerator in our sample generated before and after the two emissions caps.

Under the assumptions explained in Subsection A.4, we determine that the median incinerator observed in 1990 (before the first emissions cap) was generating 5 excess lung cancer cases per year of activity, with a latency between 5 and 15 years. In 2004, right after the second emissions cap, the same incinerator generated 0.5 excess lung cancer cases with one year of activity, with a latency of 10-15 years. We conclude that over time the health consequences of the activity of the median incinerator have decreased by a factor of 10. Still, the long-term impact of its environmental externalities accounts for around 40% of its revenues.

8 Conclusions

Pinning down the long-term effects of pollution on human health is of critical policy importance but challenging to do empirically. In this paper, we have tackled some of the main issues (minimizing measurement error to reconstruct exposure history at the individual level, addressing the endogeneity of location, and studying the interplay among cumulative exposure, intensity, and latency) and provided estimates that can have a causal interpretation.

As a source of variation, we have used heterogeneity among citizens of Danish municipalities in exposure to emissions produced by waste incinerators. Because these data cover a whole

population, they are particularly apt to identify effects on small probability outcomes, as is the case with lung cancer. Because in Denmark individual-level data on socioeconomic and health status are available for decades, it has been possible for us to trace the long-term effects on lung cancer risk of exposure to pollution.

Openings of new incinerators in municipalities that had none have allowed us to document that lung incidence increases with a decade-long delay since the start of exposure to emissions. Closings of incinerators have allowed us to identify latency and document that cumulative exposure, intensity, and latency interact in meaningful and plausible ways: latency is shorter when intensity is higher, keeping cumulative exposure constant.

The presence of long delays and interactions with different dimensions of exposure matter for policy evaluation and regulation design. Thanks to our estimates, we have shown that overlooking latency can lead to underestimate the damages of pollution and to reach wrong conclusions about the benefits of environmental regulation on health. In future work, in line with the exercise that we have performed to quantify the decline in the health consequences of emissions from incinerators over time and as a result of emissions caps, our estimates of the effects of cumulative exposure, intensity, and latency can be used to calibrate models of emissions and of people's exposure to them. The latter can be used for the design of optimal regulatory policies and to assess the welfare benefits of the regulatory interventions.

References

Amore, M. D., Bennedsen, M., Larsen, B., and Rosenbaum, P. (2019). CEO Education and Corporate Environmental Footprint. *Journal of Environmental Economics and Management*, 94:254–273.

Anderson, M. L. (2020). As the Wind Blows: The Effects of Long-term Exposure to Air Pollution on Mortality. *Journal of the European Economic Association*, 18(4):1886–1927.

Banzhaf, H. S. and Walsh, R. P. (2008). Do People Vote with Their Feet? An Empirical Test of Tiebout. *American Economic Review*, 98(3):843–63.

Banzhaf, S., Ma, L., and Timmins, C. (2019). Environmental Justice: The Economics of Race, Place, and Pollution. *Journal of Economic Perspectives*, 33(1):185–208.

Barreca, A. I., Neidell, M., and Sanders, N. J. (2021). Long-run Pollution Exposure and Mortality: Evidence from the Acid Rain Program. *Journal of Public Economics*, 200:104440.

Buekens, A. (2013). Incineration Technologies. Springer: New York, 2013.

Chay, K. Y. and Greenstone, M. (2003). The Impact of Air pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics*, 118(3):1121–1167.

Chen, Y., Ebenstein, A., Greenstone, M., and Li, H. (2013). Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences*, 110(32):12936–12941.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, 315(16):1750–1766.

Christensen, P., Sarmiento-Barbieri, I., and Timmins, C. (2022). Housing Discrimination and the Toxics Exposure Gap in the United States: Evidence from the Rental Market. *Review of Economics and Statistics*, 104(4):807–818.

Currie, J., Davis, L., Greenstone, M., and Walker, R. (2015). Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant openings and Closings. *American Economic Review*, 105(2):678–709.

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

Currie, J., Neidell, M., and Schmieder, J. F. (2009). Air Pollution and Infant Health: Lessons from New Jersey. *Journal of Health Economics*, 28(3):688–703.

Currie, J. and Schmieder, J. F. (2009). Fetal Exposures to Toxic Releases and Infant Health. *American Economic Review*, 99(2):177–183.

Currie, J. and Walker, R. (2011). Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1):65–90.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019). The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *American Economic Review*, 109(12):4178–4219.

Ebenstein, A., Fan, M., Greenstone, M., He, G., and Zhou, M. (2017). New Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences*, 114(39):10384–10389.

Filigrana, P., Milando, C., Batterman, S., Levy, J. I., Mukherjee, B., Pedde, M., Szpiro, A. A., and Adar, S. D. (2022). Exposure to Primary Air Pollutants Generated by Highway Traffic and Daily Mortality Risk in Near-Road Communities: A Case-Crossover Study. *American Journal of Epidemiology*, 191(1):63–74.

Graff Zivin, J. and Neidell, M. (2013). Environment, Health, and Human Capital. *Journal of Economic Literature*, 51(3):689–730.

Grainger, C. and Ruangmas, T. (2018). Who Wins from Emissions Trading? Evidence from California. *Environmental and Resource Economics*, 71(3):703–727.

Henneman, L. R., Choirat, C., and Zigler, C. M. (2019). Accountability Assessment of Health Improvements in the United States Associated with Reduced Coal Emissions between 2005 and 2012. *Epidemiology*, 30(4):477.

Hernandez-Cortes, D. and Meng, K. C. (2020). Do Environmental Markets Cause Environmental Injustice? Evidence from California's Carbon Market. *NBER Working Paper*.

Holgaard, J. E. and Jørgensen, T. H. (2005). A Decade of Mandatory Environmental Reporting in Denmark. *European Environment*, 15(6):362–373.

Hystad, P., Demers, P. A., Johnson, K. C., Carpiano, R. M., and Brauer, M. (2013). Long-Term Residential Exposure to Air Pollution and Lung Cancer Risk. *Epidemiology*, pages 762–772.

Kleis, A. and Dalager, S. (2007). 100 Years of Incineration in Denmark. Babcock & Wilcox Vølund A/S and Rambøll. Copenhagen: 2007.

Krause, M. and Smith, S. (2006). Final Methodology to Calculate Particulate Matter (PM) 2.5 and PM 2.5 Significance Threshold. *Diamond Bar, CA: South Coast Air Quality Management District.*

Krewski, D., Jerrett, M., Burnett, R. T., Ma, R., Hughes, E., Shi, Y., Turner, M. C., Pope III, C. A., Thurston, G., Calle, E. E., et al. (2009). *Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality*. Health Research Institute, Report 140.

Laden, F., Schwartz, J., Speizer, F. E., and Dockery, D. W. (2006). Reduction in Fine Particulate Air Pollution and Mortality: Extended Follow-up of the Harvard Six Cities sStudy. *American Journal of Respiratory and Critical Care Medicine*, 173(6):667–672.

Lepeule, J., Laden, F., Dockery, D., and Schwartz, J. (2012). Chronic Exposure to Fine Particles and Mortality: An Extended Follow-Up of the Harvard Six Cities Study from 1974 to 2009. *Environmental Health Perspectives*, 120(7):965–970.

Lubin, J. H., Caporaso, N., Wichmann, H. E., Schaffrath-Rosario, A., and Alavanja, M. C. (2007).

Cigarette Smoking and Lung Cancer: Modeling Effect Modification of Total Exposure and Intensity. *Epidemiology*, pages 639–648.

Moretti, E. and Neidell, M. (2011). Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles. *Journal of Human Resources*, 46(1):154–175.

Nadler, D. L. and Zurbenko, I. G. (2014). Estimating Cancer Latency Times Using a Weibull Model. *Advances in Epidemiology*, 2014.

Næss, Ø., Nafstad, P., Aamodt, G., Claussen, B., and Rosland, P. (2007). Relation between Concentration of Air pollution and Cause-Specific Mortality: Four-Year Exposures to Nitrogen Dioxide and Particulate Matter Pollutants in 470 Neighborhoods in Oslo, Norway. *American Journal of Epidemiology*, 165(4):435–443.

Nafstad, P., Håheim, L., Oftedal, B., Gram, F., Holme, I., Hjermann, I., and Leren, P. (2003). Lung Cancer and Air Pollution: A 27 Year Follow Up of 16 209 Norwegian Men. *Thorax*, 58(12):1071–1076.

Pennington, A. F., Strickland, M. J., Klein, M., Zhai, X., Bates, J. T., Drews-Botsch, C., Hansen, C., Russell, A. G., Tolbert, P. E., and Darrow, L. A. (2018). Exposure to Mobile Source Air Pollution in Early Life and Childhood Asthma Incidence: The Kaiser Air Pollution and Pediatric Asthma Study. *Epidemiology*, 29(1):22.

Pitot, H. C. (1993). The Molecular Biology of Carcinogenesis. *Cancer*, 72(S3):962–970.

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M. J., Brunekreef, B., et al. (2013). Air Pollution and Lung Cancer Incidence in 17 European Cohorts: Prospective Analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *The Lancet Oncology*, 14(9):813–822.

Richardson, D. B. (2009). Latency Models for Analyses of Protracted Exposures. *Epidemiology*, 20(3):395.

Shin, M., Kim, O.-J., Yang, S., Choe, S.-A., and Kim, S.-Y. (2022). Different Mortality Risks

of Long-Term Exposure to Particulate Matter across Different Cancer Sites. *International Journal of Environmental Research and Public Health*, 19(6):3180.

Sullivan, D. M. (2017). The True Cost of Air Pollution: Evidence from the Housing Market. *Working Paper*.

Swanton, C. (2023). An Actionable Inflammatory Axis for Air Pollution Indcuded Non-Small Cell Lung Cancer. *Unpublished Manuscript*.

Thy, C. (2003). The Danish Green Accounts: Experiences and Internal Effects. In: Environmental Management Accounting: Purpose and Progress. Bennet, M., Rickhardsson, P., and Shaltagger, S.. Springer, Dordrecht: 2003.

Weibust, I. (2005). A (Slow) Burning Issue: Convergence in National Regulation of Dioxins from Incineration. *Policy and Society*, 24(2):46–73.

Weisburger, J. H. (1990). The Mechanism of Lung Carcinogenesis and Smoking Cessation. *Epidemiology*, 1(4):314–317.

Figures

Figure 1: Differences in Lung Cancer Incidence Between Exposed and Unexposed Individuals Observed Before the Opening of an Incinerator Across Municipalities. This figure is drawn after running a linear probability model in which the dependent variable is lung cancer and the explanatory variables are dummies for pair of municipalities as well as for the interactions between the former and a dummy that identifies the exposed individuals. The sample is built by considering the openings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year. Each municipality involved in an opening is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious opening year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. The regression is run by restricting the attention to the period before the opening year, actual of fictitious. The sample size is 1,498,472 person-years, with the exposed accounting for 918,145 observations. The figure shows (in basis points) the coefficients on the interactions between the pairs of municipalities and the dummy identifying the exposed, together with the 95% confidence intervals. Each coefficient is labeled with the name of the municipality where the actual opening occurs.

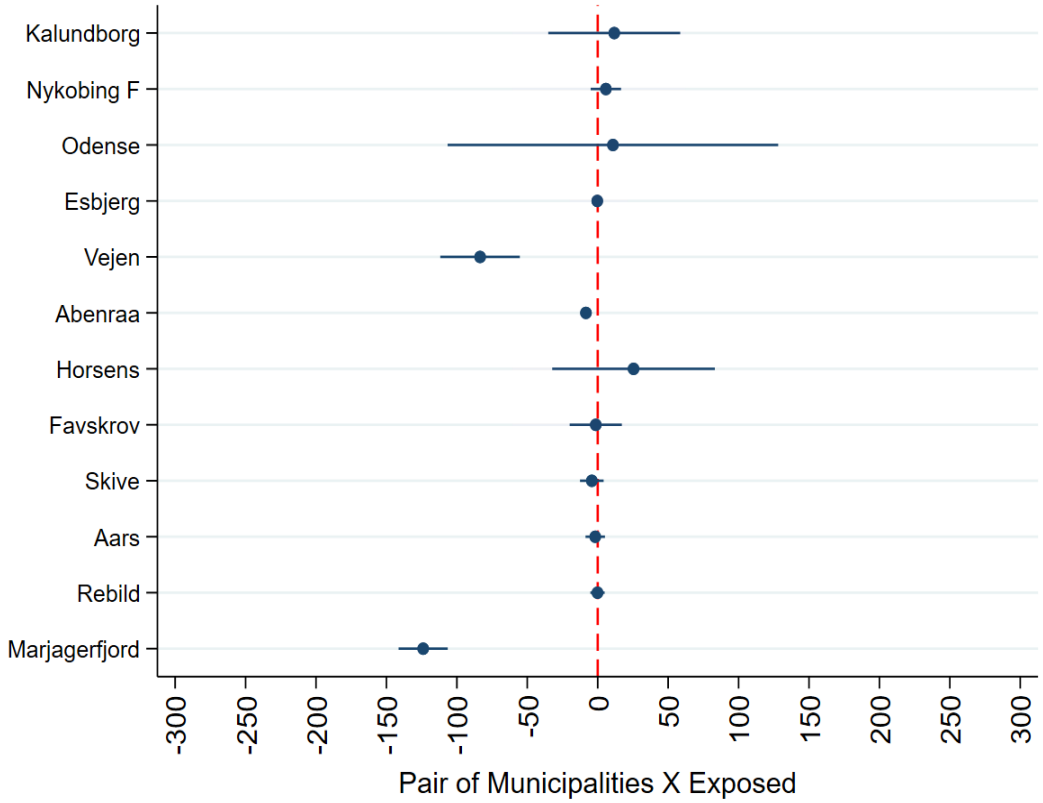


Figure 2: Differences in Lung Cancer Incidence Between Exposed and Unexposed Individuals Observed Before the Closing of an Incinerator Across Municipalities. This figure is drawn after running a linear probability model in which the dependent variable is lung cancer and the explanatory variables are dummies for pair of municipalities as well as for the interactions between the former and a dummy that identifies the exposed individuals. The sample is built by considering the closings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was shut down, before or during the closing year. Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious closing year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. The regression is run by restricting the attention to the period before the closing year, actual of fictitious. The sample size is 4,288,021 person-years, with the exposed accounting for 2,863,583 observations. The figure shows (in basis points) the coefficients on the interactions between the pairs of municipalities and the dummy identifying the exposed, together with the 95% confidence intervals. Each coefficient is labeled with the name of the municipality where the actual closing occurs. We omit one coefficient (related to Solrod), which has a very wide confidence interval, due to lack of statistical power; the coefficient is positive but not statistically significant.

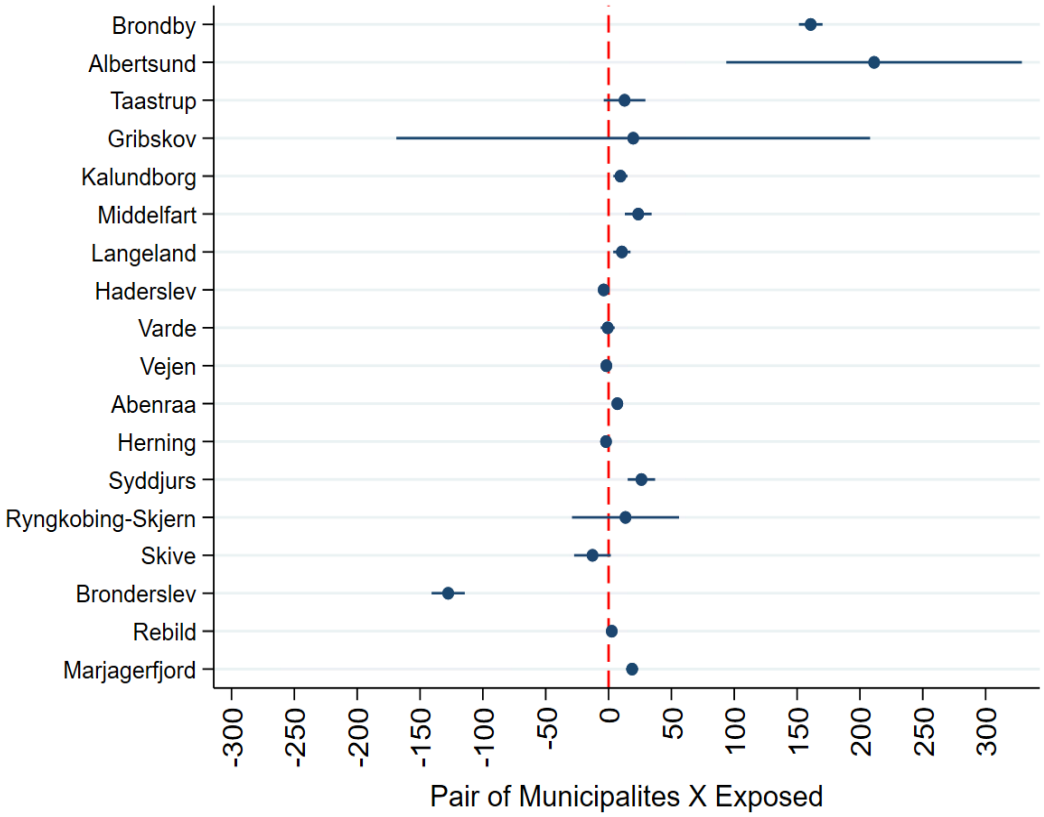


Figure 3: Share of Movers as a Function of Time Since Opening. This picture shows the average of a dummy that takes the value of 1 with reference to person-year i-t if i is associated with two different parishes in $t+1$ and t, as a function of years since the opening of an incinerator and distinguishing between exposed and unexposed. The sample is built by considering the openings of incinerators between 1982 and 2018 (when we have the full set of controls). The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year, when they are first observed between 1975 and 2018 (when we have geocoded addresses). Each municipality involved in an opening is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious opening year that depends on the matched municipality, when they are first observed between 1975 and 2018. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. If an individual changes parish during his/her lifetime, we consider only the person-years prior to the change in parish. The sample size is 10,532,589 person-years observed between 1982 and 2018.

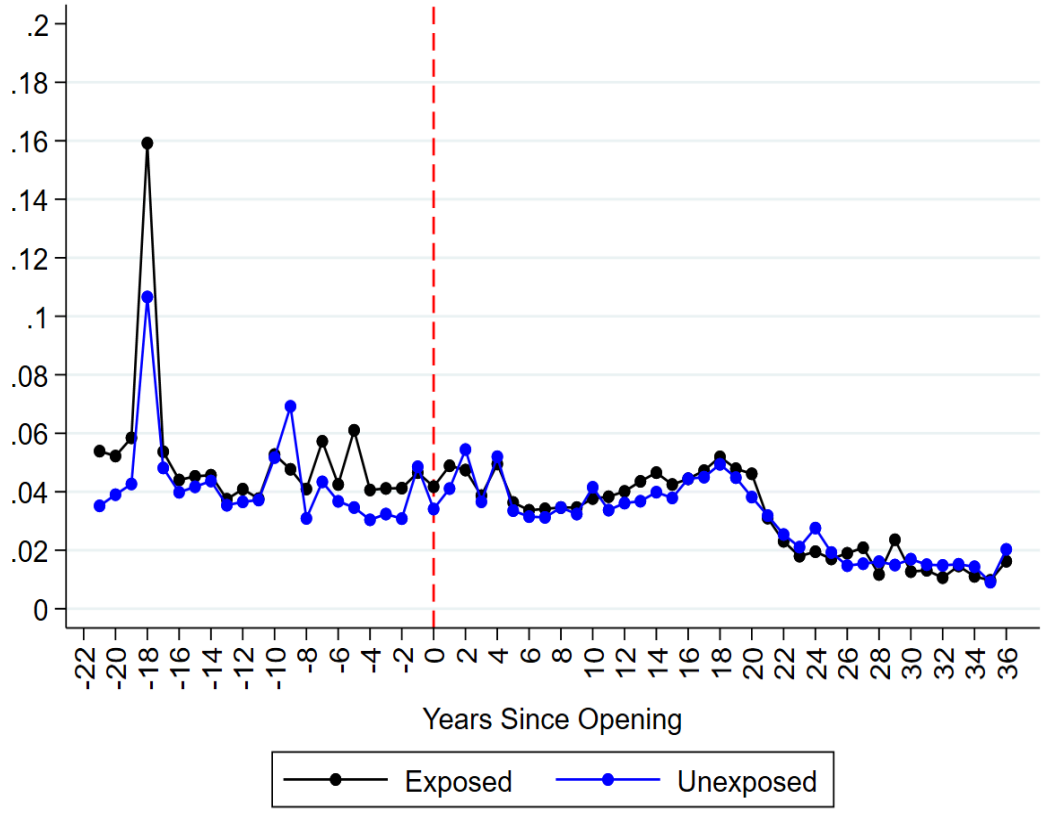


Figure 4: Share of Movers as a Function of Time Since Closing. This picture shows the average of a dummy that takes the value of 1 with reference to person-year i-t if i is associated with two different parishes in $t+1$ and t, as a function of years since the closing of an incinerator and distinguishing between exposed and unexposed. The sample is built by considering the closings of incinerators between 1982 and 2018 (when we have the full set of controls). The exposed are identified as individuals living in a municipality where an incinerator was shut down, before or during the closing year, when they are first observed between 1975 and 2018 (when we have geocoded addresses). Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious closing year that depends on the matched municipality, when they are first observed between 1975 and 2018. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. If an individual changes parish during his/her lifetime, we consider only the person-years prior to the change in parish. The sample size is 12,555,605 person-years observed between 1982 and 2018.

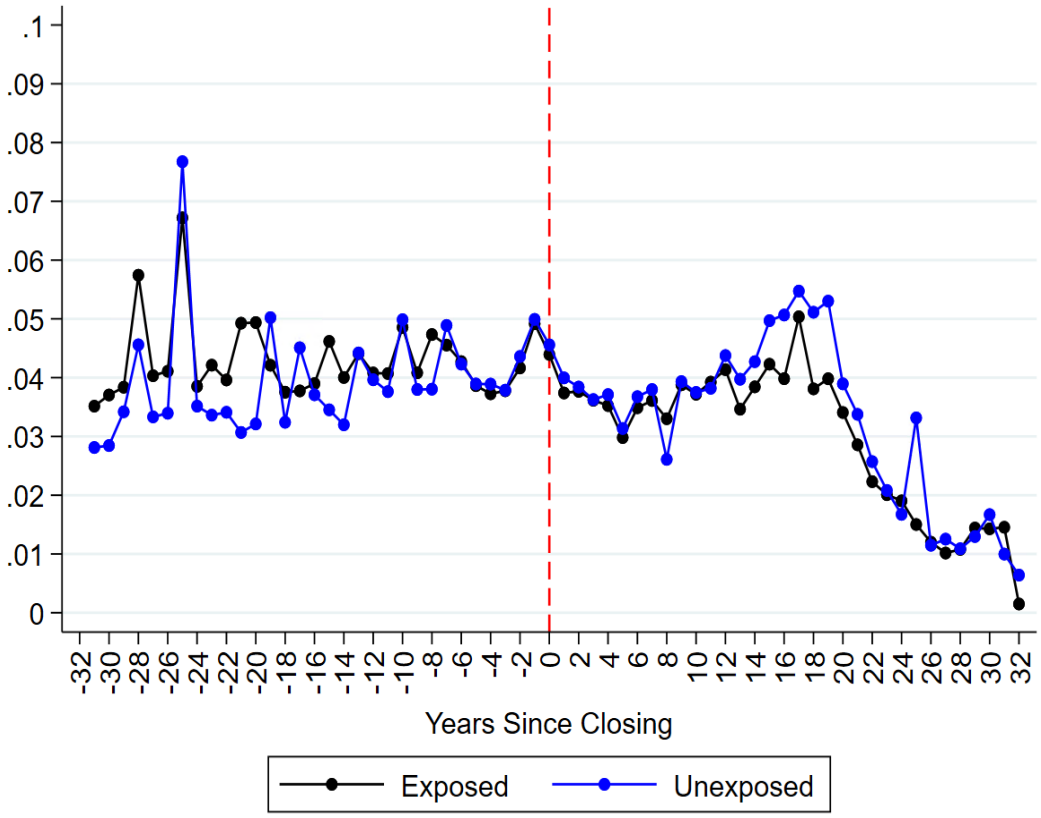


Figure 5: Actual and Fitted Emissions of SO2. This figure shows the total amount of SO2 (in tonnes) emitted from incinerators between 1975 and 2018, distinguishing between actual and fitted emissions. Actual emissions are drawn from the Green Accounts available for the periods 1996-2001 and 2007-2018 (including some observations resulting from historical data about previous years and presented in the reports related to these two periods). Fitted emissions are computed as the fitted values from a regression in which actual emissions of SO2 are predicted as a function of imputed emissions of SO2 and concentrations of SO2 at the parish level. Imputed emissions are computed as the product between emissions factors (from the CORINAIR Program) and amount of waste burnt (from the Danish Energy Agency); for the period prior to 1990, we assume that the 1990 emissions factors stayed constant; for the period prior to 1994, we discount the 1994 data on waste burnt by real GDP growth. Concentrations are downloaded from the MERRA-2 dataset provided by the NASA; for the period prior to 1980, we assume that the 1980 concentrations stayed constant; the 0.5^{o}x0.625^{o} resolution of the NASA is matched with the shapefile of the Danish parishes.

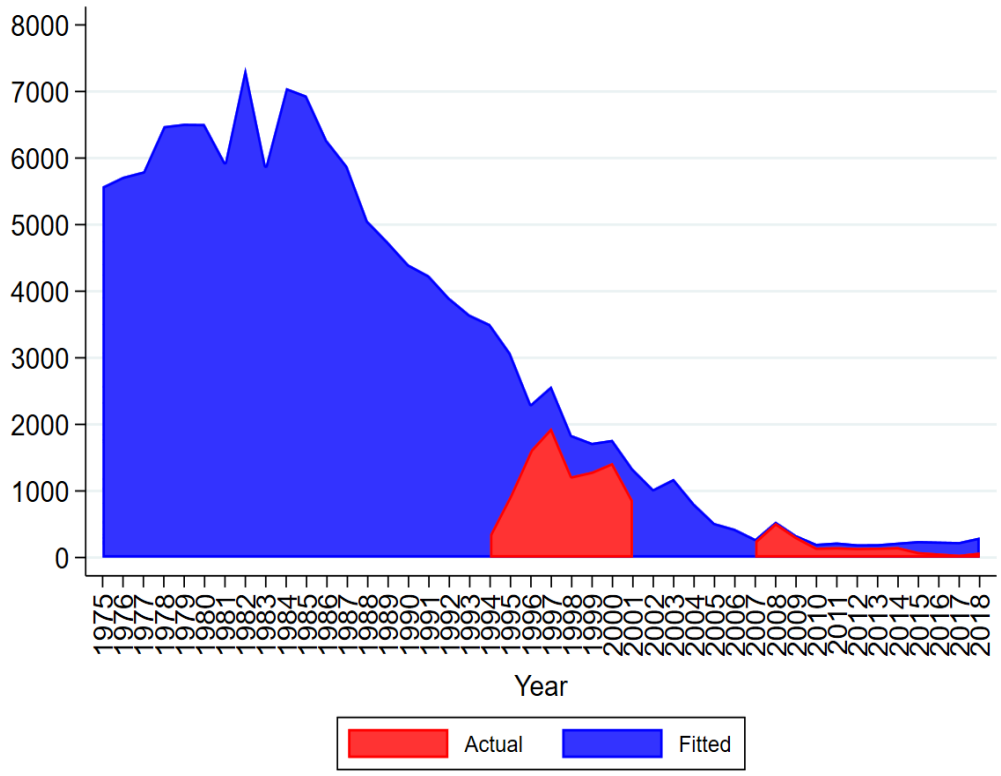
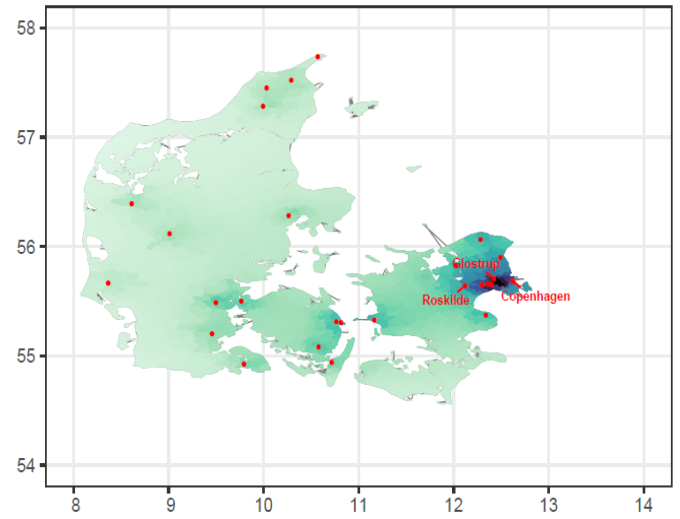
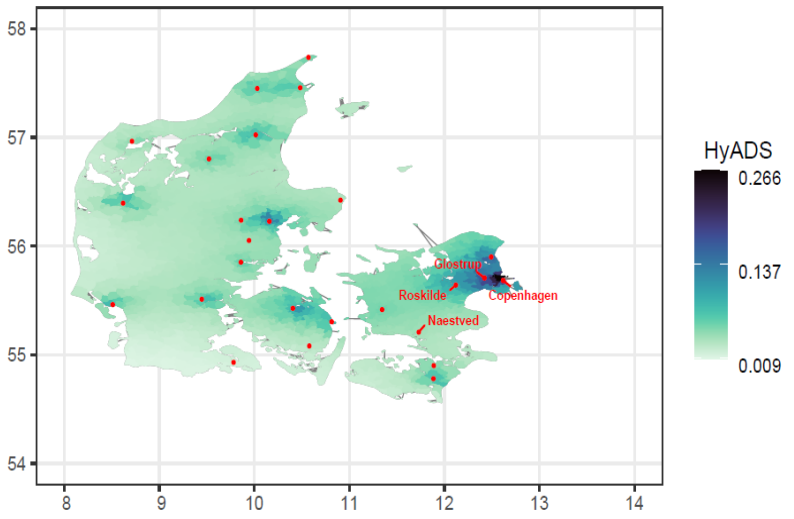


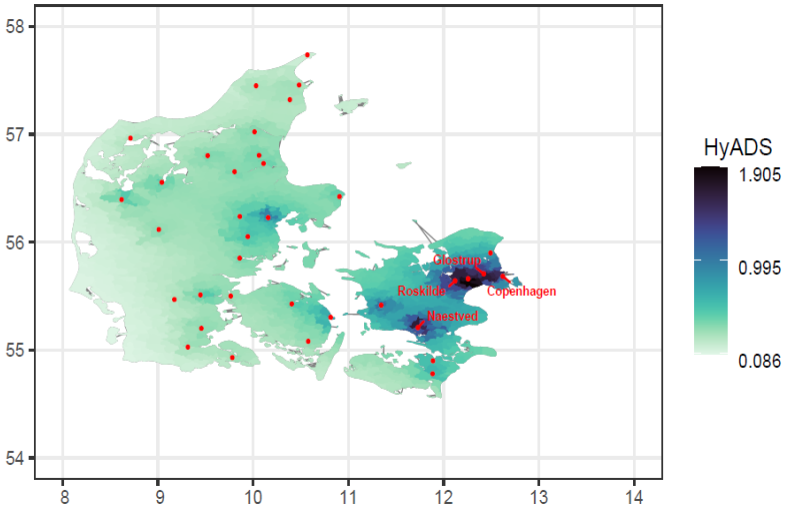
Figure 6: HyADS in 1975, 1997, and 2018. This figure shows the map of Denmark in 1975, 1997, and 2018, with each parish being colored depending on the value taken by HyADS and each red dot representing the location of an active incinerator. HyADS is computed by running HYSPLIT starting from emissions of SO2 from incinerators; every year, the release of 100 particles is simulated from each active plant, 4 times per day; the concentration of the particles from each plant is computed at the parish level and weighted by emissions of SO2. The latter are either drawn from the Green Accounts (available in 1996-2001 and in 2007-2018) or computed as the fitted values from a regression in which actual emissions of SO2 are predicted as a function of imputed emissions of SO2 and concentrations of SO2. Imputed emissions are computed as the product between emissions factors (from the CORINAIR Program) and amount of waste burnt (from the Danish Energy Agency); for the period prior to 1990, we assume that the 1990 emissions factors stayed constant; for the period prior to 1994, we discount the 1994 data on waste burnt by real GDP growth. Concentrations are downloaded from the MERRA-2 dataset provided by the NASA; for the period prior to 1980, we assume that the 1980 concentrations stayed constant; the 0.5^{o}x0.625^{o} resolution of the NASA is matched with the shapefile of the Danish parishes.



(a) HyADS in 1975.

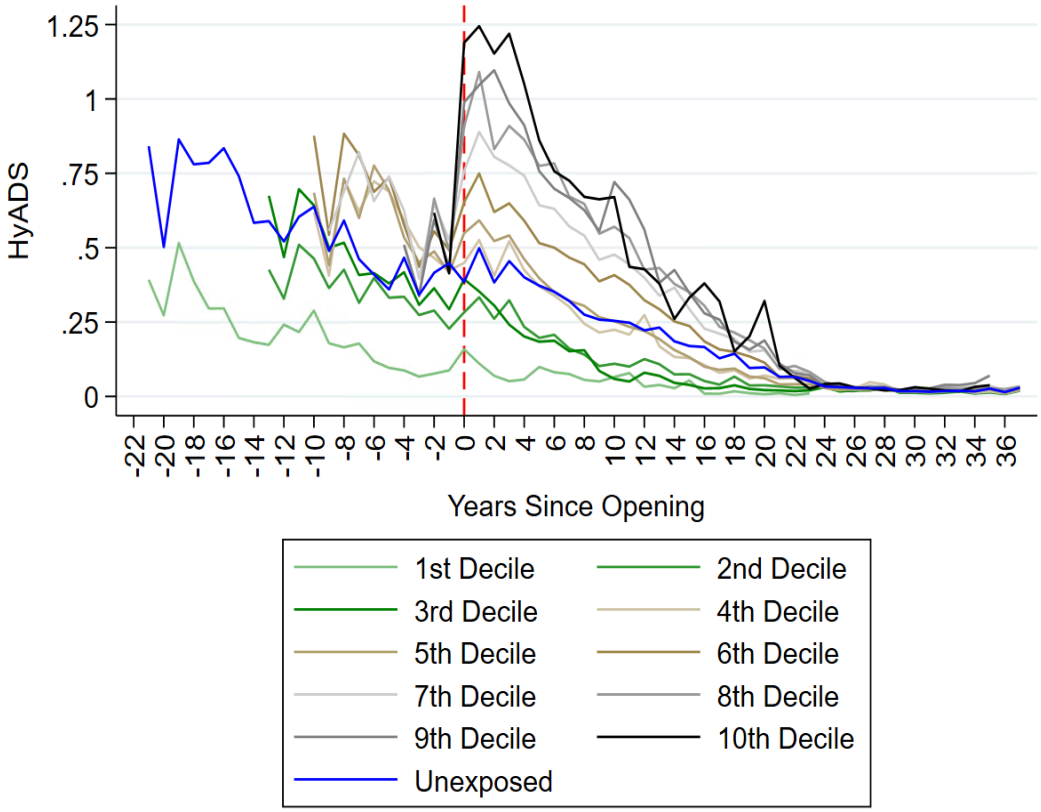


(b) HyADS in 2018.



(c) HyADS in 1997.

Figure 7: Average HyADS for the Exposed (Divided by Decile) and the Unexposed in the Sample of Openings. This picture shows the average number of HyADS units as a function of years since opening, distinguishing between exposed individuals (divided into deciles according to the distribution of HyADS at the time of the opening) and unexposed individuals. The sample is built by considering the openings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year. Each municipality involved in an opening is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious opening year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime.



Tables

Table 1: T-Tests Before the Opening of an Incinerator. This table shows the results for T-tests related to the control variables (age, household income in 2015-DKK, household wealth in 2015-DKK, sex, and marital status) in the sample of openings. Such tests are run by considering the period before the opening year, actual or fictitious. The sample is built by considering the openings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year. Each municipality involved in an opening is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious opening year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

Variable	Exposed	Unexposed	Difference
Age	54.569	55.820	-1.251***
Income	182,478	176,154	6,324***
Wealth	545,936	543,703	2,223
Male (%)	49.795	49.716	0.080
Married (%)	68.948	66.359	2.589***
Divorced (%)	3.429	3.707	-0.278***
Widower (%)	13.767	15.230	-1.462***
N	918,145	580,327	

Table 2: T-Tests Before the Closing of an Incinerator. This table shows the results for T-tests related to the control variable (age, household income in 2015-DKK, household wealth in 2015-DKK, sex, and marital status) in the sample of closings. Such tests are run by considering the period before the closing year, actual of fictitious. The sample is built by considering the closings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was shut down, before or during the closing year. Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious closing year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

Variable	**Exposed**	**Unexposed**	**Difference**
Age	53.038	51.595	1.443***
Income	191,922	188,300	3,622***
Wealth	649,645	664,493	-14,848***
Male (%)	50.524	51.318	-0.794***
Married (%)	64.036	60.906	3.129***
Divorced (%)	3.419	2.897	0.521***
Widower (%)	13.009	13.009	-0.0006
N	2,863,583	1,424,438	

Table 3: T-Tests Within 5 Years Before and Within 5 Years After the Openings for Mover-Years. This table shows the results for T-tests related to the control variables (age, household income in 2015-DKK, household wealth in 2015-DKK, sex, and marital status) in the sample of openings, restricting the attention to the movers observed the year before the change in parish. Such tests are run both within 5 years before and within 5 years after the opening. The sample is built by considering the openings of incinerators between 1982 and 2018 (when we have the full set of controls). The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year, when they are first observed between 1975 and 2018 (when we have geocoded addresses). Each municipality involved in an opening is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious opening year that depends on the matched municipality, when they are first observed between 1975 and 2018. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. In this table, we restrict the attention to people who change their parish either within 5 years before or within 5 years after the opening and we keep only the observation prior to the change in parish for each individual. *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

	5 Years Before			**5 Years After**		
Variable	**Exposed**	**Unexposed**	**Difference**	**Exposed**	**Unexposed**	**Difference**
Age	27.696	31.165	-3.468***	29.271	32.309	-3.037***
Income	181,159	179,313	1,846*	186,385	184,247	2,138
Wealth	363,782	369,068	-5,826	433,168	464,046	-30,879***
Male (%)	49.243	49.977	-0.734	49.617	49.013	0.604*
Married (%)	22.840	26.967	-4.127***	24.307	30.302	-5.995***
Divorced (%)	2.146	2.470	-0.324**	2.247	2.262	-0.015
Widower (%)	5.754	7.550	-1.796***	6.022	7.043	-1.022***
N	35,924	17,206		55,084	40,179	

Table 4: T-Tests Within 5 Years Before and Within 5 Years After the Closings for Mover-Years. This table shows the results for T-tests related to the control variables (age, household income in 2015-DKK, household wealth in 2015-DKK, sex, and marital status) in the sample of closings, restricting the attention to the movers observed the year before the change in parish. Such tests are run both within 5 years before and within 5 years after the opening. The sample is built by considering the closings of incinerators between 1982 and 2018 (when we have the full set of controls). The exposed are identified as individuals living in a municipality where an incinerator was shut down, before or during the closing year, when they are first observed between 1975 and 2018 (when we have geocoded addresses). Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious closing year that depends on the matched municipality, when they are first observed between 1975 and 2018. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. In this table, we restrict the attention to people who change their parish either within 5 years before or within 5 years after the closing and we keep only the observation prior to the change in parish for each individual. *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

	5 Years Before			**5 Years After**		
Variable	**Exposed**	**Unexposed**	**Difference**	**Exposed**	**Unexposed**	**Difference**
Age	29.457	26.851	2.607***	29.045	25.470	3.575***
Income	200,398	193,272	7,126***	207,451	199,014	8,438***
Wealth	536,879	592,418	-55,540***	574,414	556,341	18,072
Male (%)	49.645	49.574	0.071	49.844	49.658	0.185
Married (%)	22.086	18.689	3.396***	18.669	15.667	3.002***
Divorced (%)	2.095	1.444	0.651***	1.767	1.453	0.314***
Widower (%)	6.349	5.770	0.578***	6.436	5.181	1.255***
N	49,018	27,347		40,629	25,321	

Table 5: Prediction Model for Actual SO2 Emissions from the Green Accounts. This table shows the results for OLS specifications in which the dependent variable is the amount of SO2 (in tonnes) emitted by an incinerator in a given year according to the Green Accounts. The explanatory variables are imputed SO2 emissions (in tonnes) and SO2 concentration (in μg/m^3) at the parish level. Imputed emissions are computed as the product between emissions factors (from the CORINAIR Program) and amount of waste burnt (from the Danish Energy Agency). Concentrations are from the MERRA-2 dataset provided by the NASA; the 0.5^{o}x0.625^o resolution of the NASA is matched with the shapefile of the Danish parishes; the attention is restricted to the parishes associated with the relevant incinerators. The sample consists of the incinerator-years that report SO2 emissions on the Green Accounts for the periods 1996-2001 and 2007-2018 (451 observations, including some observations resulting from historical data about previous years and presented in the reports related to these two periods). Standard errors are in parentheses; *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

Actual SO2 Emissions	(I)	(II)	(III)
Imputed SO2 Emissions	1.067***	-	0.815***
	(0.067)		(0.074)
SO2 NASA	-	22.792***	12.523***
		(1.805)	(1.858)
Constant	6.872***	-55.451***	-33.834***
	(2.407)	(6.922)	(6.461)
R^2	0.359	0.262	0.418
N	451	451	451

Table 6: SO2 Concentration from Monitors as a Function of HyADS Based on Actual and Fitted Emissions of SO2. This table shows the results for OLS specifications in which the dependent variable is the annual average concentration of SO2 (in μg/m^3) measured by the monitors of the University of Aarhus. The main explanatory variable is HyADS based on emissions of SO2 from incinerators. Each monitor is matched with a parish depending on its coordinates. HyADS is computed by running HYSPLIT; every year, the release of 100 particles is simulated from each active plant, 4 times per day; the concentration of the particles from each plant is computed at the parish level and weighted by emissions of SO2. The latter are either drawn from the Green Accounts (available in 1996-2001 and in 2007-2018) or computed as the fitted values from a regression in which actual emissions of SO2 are predicted as a function of imputed emissions of SO2 and concentrations of SO2. Imputed emissions are computed as the product between emissions factors (from the CORINAIR Program) and amount of waste burnt (from the Danish Energy Agency); for the period prior to 1990, we assume that the 1990 emissions factors stayed constant; for the period prior to 1994, we discount the 1994 data on waste burnt by real GDP growth. Concentrations are downloaded from the MERRA-2 dataset provided by the NASA; for the period prior to 1980, we assume that the 1980 concentrations stayed constant; the 0.5^{o}x0.625^o resolution of the NASA is matched with the shapefile of the Danish parishes. We control for year and monitor fixed effects. Standard errors are in parentheses; *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

SO2 Monitors	(I)	(II)	(III)
SO2 HyADS	6.395***	3.481***	4.467***
	(0.592)	(0.253)	(0.284)
Constant	2.494***	4.966***	19.019***
	(0.748)	(0.291)	(0.898)
Year FE	No	Yes	Yes
Monitor FE	No	No	Yes
R^2	0.526	0.961	0.986
N	107	107	107

Table 7: Distribution of Cumulative HyADS. This table shows the distribution of cumulative HyADS with reference to the sample consisting of the Danish people aged at least 2 observed between 1982 and 2018; we drop the people without information on residential address for at least one year and the people who spend at least one year in a Municipality associated with an incinerator for which we are not able to identify opening and closing years (Helsingor, Svejstrup, Skibby, Struer, and Ringsted). The sample size is 169,920,085 person-years.

Percentile	Cum. HyADS
0%	0.005
1%	0.182
5%	0.943
10%	2.339
25%	7.415
50%	15.500
75%	32.097
90%	74.213
95%	96.636
99%	131.404
100%	157.560

Table 8: Temporal Distribution of Openings and Closings. This table shows the temporal distribution of the openings and closings of the incinerators considered to run the specifications fitted to the samples of openings and closings respectively. In order for an opening to be included, it must occur in a municipality where an incinerator has never operated before. In order for a closing to be included, it must occur in a municipality where no incinerator will be opened afterwards.

Year	Openings	Closings
1982	2	
1983	2	
1984	2	
1985		2
1986	2	1
1987		1
1990	1	1
1991	1	
1992	2	2
1993		1
1995	1	2
1996	1	
1998		1
2000		1
2003	1	
2005		1
2006		1
2010		1
2013		1
2017		1
Total	15	20

Table 9: Median Opening Year for the Incinerators Associated with Different Deciles in the Sample of Openings. This table shows the median of the opening year of the incinerators associated with different deciles (identified by considering the distribution of HyADS at the time of the opening) among the exposed individuals in the sample of openings. The sample is built by considering the openings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year year. The median opening year is determined by considering the distribution across the unique incinerators associated with the individuals included in a given decile (not across the individuals).

Decile	Median Opening Year
1st	1995
2nd	1995
3rd	1995
4th	1986
5th	1986
6th	1984
7th	1984
8th	1984
9th	1984
10th	1984

Table 10: Lung Cancer as a Function of Time Since Opening. This table shows the results for linear probability models in which the dependent variable is lung cancer. The main explanatory variables are categories of years since the opening of an incinerator for the exposed individuals. The sample is built by considering the openings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was opened, before or during the opening year. Each municipality involved in an opening is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious opening year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units before the opening year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. We control for household income and wealth quintiles lagged by two years, sex, marital status, and age range. We include municipality, pair of municipalities, and year fixed effects. Coefficients and standard errors (in parentheses) are in basis points, with the baseline probability of lung cancer being 23 basis points. *, **, and *** mean statistically significant at the 10, 5, and 1% levels. The same regression is fitted to the entire sample of openings, to the sample of the bottom 3 deciles among the exposed, to the sample of the middle 4 deciles among the exposed, and to the sample of the top 3 deciles among the exposed; exposed individuals are divided into deciles according to the distribution of HyADS at the time of the opening, with the relevant cutoffs being 0.4 and 0.9 HyADS units for the categorization implemented in this table; when we restrict the attention to a subset of exposed individuals, the control group is formed by the unexposed individuals associated with the municipalities matched with those to which the relevant exposed individuals belong.

	(I)	(II)	(III)	(IV)
Years Since Opening	**Entire Sample**	**Bottom 30% Exposed**	**Middle 40% Exposed**	**Top 30% Exposed**
0-4 Years	0.779	2.117	-0.291	6.444**
	(0.915)	(1.468)	(1.498)	(2.902)
5-9 Years	0.070	-1.587	1.563	6.934**
	(0.991)	(1.579)	(1.630)	(3.062)
10-14 Years	1.339	-2.650	3.825**	10.153***
	(1.086)	(1.732)	(1.763)	(3.149)
15-19 Years	1.258	-1.891	3.768**	9.202***
	(1.241)	(2.209)	(1.852)	(3.247)
20-24 Years	1.094	3.489	0.013	7.930**
	(1.388)	(2.669)	(1.940)	(3.361)
25-29 Years	4.011**	3.504	2.426	10.439***
	(1.615)	(4.222)	(2.127)	(3.490)
30+ Years	4.822***	12.214**	4.722**	3.596
	(1.797)	(4.974)	(2.350)	(3.674)
Lagged Income Quintiles	Yes	Yes	Yes	Yes
Lagged Wealth Quintiles	Yes	Yes	Yes	Yes
Sex	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes
Age Range	Yes	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes	Yes
Pair of municipalities FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	6,144,782	1,853,088	3,203,880	2,510,719
R^2	0.001	0.001	0.001	0.002

Table 11: Hazard Ratio for 1 More 0.9-HyADS-Year as a Function of Time Since Closing. This table shows the results for a Cox regression in which the failure event is represented by lung cancer and the main explanatory variable is the amount of HyADS accumulated by an exposed individual from the most recent between birth and 1950 to the closing year of the incinerator associated with his/her municipality; this amount is divided by 0.9 and measured in 0.9-HyADS-Years; this variable is interacted with categories of years since closing. The sample is built by considering the closings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was shut down, during the closing year. Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, during the fictitious closing year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. We consider both the exposed and the unexposed as under observation since the closing year (actual or fictitious). We control for household income and wealth quintiles lagged by two years as well as for marital status. We consider age as the time scale. We stratify the sample by sex, birth cohort, and pair of municipalities. The table shows the hazard ratios for the explanatory variables of interest; each hazard ratio must be interpreted as the gross excess risk associated with 1 more 0.9-HyADS-Year of cumulative exposure before the closing, accounting for the number of years elapsed since the closing; standard errors are in parentheses; *, **, and *** mean statistically different from 1 at the 10, 5, and 1% levels.

Years Since Closing	**HR Per 0.9-HyADS-Year**
0-4 Years	1.0050**
	(0.0020)
5-9 Years	1.0052**
	(0.0021)
10-14 Years	1.0053*
	(0.0027)
15+ Years	1.0019
	(0.0033)
N	3,009,360

Table 12: Hazard Ratio for 1 More 0.9-HyADS-Year as a Function of Time Since Closing and Average Intensity of Exposure Before the Closing. This table shows the results for a Cox regression in which the failure event is represented by lung cancer and the main explanatory variable is the amount of HyADS accumulated by an exposed individual from the most recent between birth and 1950 to the closing year of the incinerator associated with his/her municipality; this amount is divided by 0.9 and measured in 0.9-HyADS-Years; this variable is interacted with different combinations of categories of years since closing and categories of the average number of HyADS units per year; the latter is obtained by dividing the amount of HyADS accumulated before the closing by the minimum of age at closing and the difference between the closing year and 1950. The sample is built by considering the closings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was shut down, during the closing year. Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, during the fictitious closing year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. We consider both the exposed and the unexposed as under observation since the closing year (actual or fictitious). We control for household income and wealth quintiles lagged by two years as well as for marital status. We consider age as the time scale. We stratify the sample by sex, birth cohort, and pair of municipalities. The table shows the hazard ratios for the explanatory variables of interest; each hazard ratio must be interpreted as the gross excess risk associated with 1 more 0.9-HyADS-Year of cumulative exposure before the closing, accounting for the number of years elapsed since the closing and for the average number of HyADS units per year; standard errors are in parentheses; *, **, and *** mean statistically different from 1 at the 10, 5, and 1% levels. For each category of intensity, the table shows the median of cumulative exposure (in 0.9-HyADS-Years), considering the distribution across the individuals included in the relevant category.

Average Intensity	Years Since Closing	HR Per 0.9-HyADS-Year	Median 0.9-HyADS-Years
0-0.4 HyADS	0-4 Years	1.0005	
		(0.0036)	
	5-10 Years	1.0037	
		(0.0043)	8.6625
	10-14 Years	1.0144**	
		(0.0065)	
	15+ Years	1.0031	
		(0.0056)	
0.4-0.9 HyADS	0-4 Years	1.0018	
		(0.0036)	
	5-10 Years	1.0068*	
		(0.0039)	24.7128
	10-14 Years	1.0106**	
		(0.0047)	
	15+ Years	1.0034	
		(0.0042)	
$\geq$0.9 HyADS	0-4 Years	1.0074***	
		(0.0024)	
	5-10 Years	1.0057**	
		(0.0026)	81.0938
	10-14 Years	1.0031	
		(0.0032)	
	15+ Years	1.0017	
		(0.0044)	
	N	3,009,360	

Table 13: Smoking Habits as a Function of Exposure to Incinerators. This table shows the results for linear probability models aimed at predicting a dummy that takes the value of 1 if a pregnant woman is classified as a current smoker. The sample consists of pregnant-woman-years observed between 1999 and 2018. The explanatory variables are HyADS at the Parish-year level computed with reference to SO2 emissions from incinerators, log household income and wealth (in 2015-DKK), marital status, and age. We control for year and municipality fixed effects. Coefficients and standard errors (in parentheses) are in percentage points, with the baseline probability of smoking being of 15%. *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

Smoker	(I)	(II)
HyADS	-0.611**	-3.254***
	(0.247)	(0.241)
Log Income	-	-1.591***
		(0.051)
Log Wealth	-	-2.241***
		(0.014)
Married	-	-7.587***
		(0.068)
Divorced	-	6.433***
		(0.177)
Widower	-	6.408***
		(1.046)
Age	-	-0.285***
		(0.007)
Constant	23.085***	81.082***
	(0.197)	(0.635)
Year FE	Yes	Yes
Municipality FE	Yes	Yes
N	1,199,719	1,199,719
R^2	0.029	0.081

Table 14: Lung Cancer as a Function of Cumulative HyADS, Smoking Habits, and Other Controls. This table shows the results for linear probability models aimed at predicting lung cancer. The sample consists of women who have been pregnant between 1999 and 2018, observed between 1982 and 2018. The main explanatory variables are the quartiles of cumulative HyADS, computed at the Parish level with reference to SO2 emissions from incinerators. The main control is Ever Smoker, a dummy that takes the value of 1 with reference to a woman who is ever classified as a smoker between 1999 and 2018 while being pregnant; this variable is fixed with reference to the same woman. The other controls are age range, household income and wealth quintiles lagged by two years, and marital status. We also add year and municipality fixed effects. Coefficients and standard errors (in parentheses) are in basis points. The baseline probability of lung cancer is 0.33 basis points. *, **, and *** mean statistically significant at the 10, 5, and 1% levels. R^2s are in basis points.

Lung Cancer	(I)	(II)
Cum. HyADS 25-50%	0.064*	0.045
	(0.038)	(0.039)
Cum. HyADS 50-75%	0.266***	0.234***
	(0.043)	(0.043)
Cum. HyADS 75-100%	0.348***	0.298***
	(0.051)	(0.051)
Ever Smoker	-	0.519***
		(0.034)
Age <30	-0.525***	-0.513***
	(0.045)	(0.045)
Age >60	46.470***	46.472***
	(1.637)	(1.637)
Constant	0.403***	0.303**
	(0.122)	(0.122)
Lagged Income Quintiles	Yes	Yes
Lagged Wealth Quintiles	Yes	Yes
Marital Status	Yes	Yes
Year FE	Yes	Yes
Municipality FE	Yes	Yes
R^2 (bp)	1.423	1.550
N	18,854,406	18,854,406

A Appendix

A.1 Justifying the Use of Imputed Emissions and NASA Concentrations as Predictors for SO2 Emissions

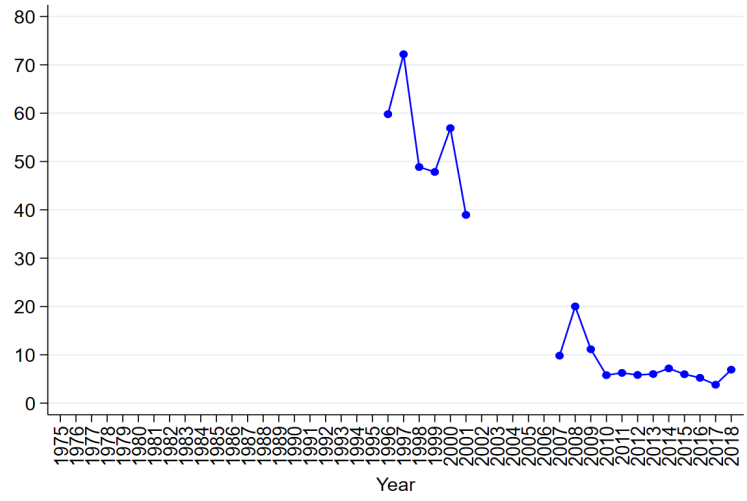
In order to assign emissions of SO2 to incinerator-years without Green Accounts, we use both imputed emissions of SO2 and NASA concentration of SO2 at the parish level. The first are computed as the product of annual sector-specfic emissions factors and annual amount of waste burnt at the plant level; as such, they are designed to proxy for actual emissions. In contrast, NASA concentrations of SO2 are related to all pollution sources and not specifically to incinerators. We justify our choice to include also NASA concentrations by resorting to a variety of reasons.

First, NASA concentrations are available from 1980, while emissions factors are available from 1990 and waste burnt from 1994. Thus, to calculate imputed emissions for the incinerator-years related to the pre-1994 period, we need to make the assumption that the 1990 emissions factors remained constant and we need to discount the 1994 amounts of waste burnt by real GDP growth. As a result, relying only on imputed emissions would render our assessment of pollution from incinerators significantly more uncertain for the first part of our sample. This argument is further supported by Figure A.1, where we plot average actual emissions as reported in the Green Accounts, average imputed emissions, and average NASA concentrations corresponding to the parishes associated with the relevant incinerators. Actual emissions follow a sharp declining trend since they start being available in 1996; imputed emissions are relatively stable between the 1980s and the early 1990s, while NASA concentrations sharply decline since the mid-1980s. Considering that pollution from the waste incineration industry was very heavy in the 1980s and that the first Executive Order capping emissions from incinerators was passed in 1991, actual emissions from incinerators in the 1980s were likely much larger than in the 1990s (see Kleis and Dalager, 2007, who provide many historical details about the waste incineration industry in Denmark and the first investigations run by the EPA in the 1980s). While SO2 concentrations from the NASA perfectly capture this pattern, imputed emissions are stable in the 1980s, due to the assumptions that we make to compute them.

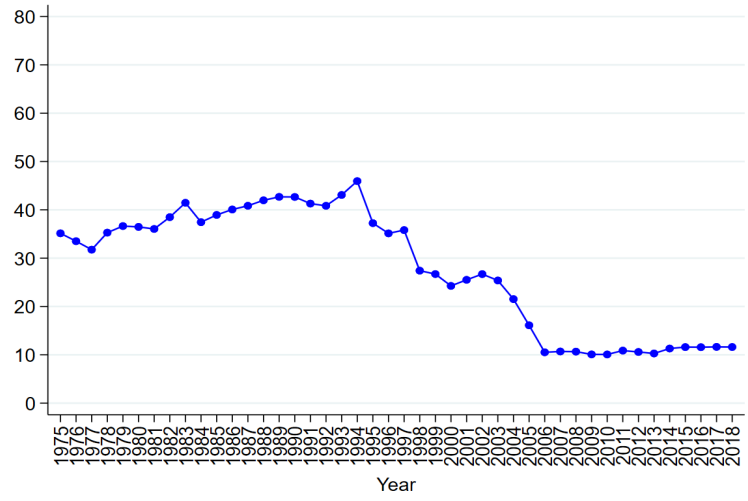
Second, the choice of using both variables is supported by the evidence that NASA concentrations of SO2 are closely linked to exposure to pollution from incinerators. Namely, in Table A.1, we regress SO2 concentrations from the NASA estimated with reference to the parishes associated with incinerators that report SO2 emissions on the Green Accounts on HyADS computed in correspondence to those parishes and by taking into account only actual emissions. The results show that HyADS alone explains more than 40% of the variance of the dependent variable.

Third, we show that SO2 concentrations from the NASA perform extremely well as a predictor of actual SO2 concentrations from monitors. Table A.2 shows that NASA concentrations alone explain more than 80% of the variance of the actual concentrations of SO2; in addition, the coefficient on the explanatory variable remains significant after controlling for year and monitor fixed effects.

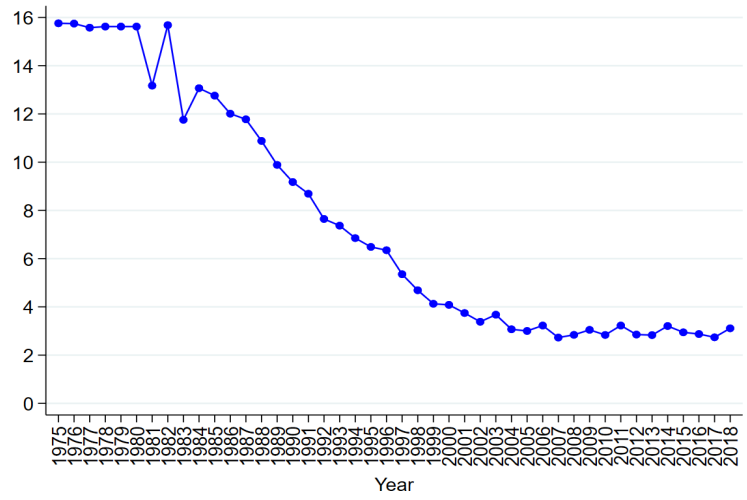
Figure A.1: Average Actual Emissions, Imputed Emissions, and Concentrations of SO2. This figure shows the annual average of actual emissions of SO2, imputed emissions of SO2, and concentrations of SO2 computed in correspondence to the parishes associated with the relevant incinerators. Actual emissions (in tonnes) are drawn from the Green Accounts for the periods 1996-2001 and 2007-2018. Imputed emissions (in tonnes) are computed as the product between emissions factors (from the CORINAIR Program) and amount of waste burnt (from the Danish Energy Agency); for the period prior to 1990, we assume that the 1990 emissions factors stayed constant; for the period prior to 1994, we discount the 1994 data on waste burnt by real GDP growth. Concentrations (in μg/m^3) are downloaded from the MERRA-2 dataset provided by the NASA; for the period prior to 1980, we assume that the 1980 concentrations stayed constant; the 0.5^{o}x0.625^o resolution of the NASA is matched with the shapefile of the Danish parishes.



(a) Actual Emissions of SO2.



(b) Imputed Emissions of SO2.



(c) NASA Concentrations of SO2.

Table A.1: SO2 Concentrations from the NASA as a Function of HyADS Based on Actual SO2 Emissions. This table shows the results for OLS specifications in which the dependent variable is the concentration of SO2 (in μg/m^3) from the MERRA-2 dataset provided by the NASA. The main explanatory variable is HyADS based on actual emissions of SO2 from incinerators as reported in the Green Accounts. The 0.5^ox0.625^o resolution of the NASA is matched with the shapefile of the Danish parishes. HyADS is obtained at the parish level by running HYSPLIT based on actual emissions of SO2 from incinerators as reported in the Green Accounts. The sample consists of the parish-years associated with the incinerator-years that report emission of SO2 in 1996-2001 and 2007-2018 (including historical data). We control for year and incinerator fixed effects. Standard errors are in parentheses; *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

SO2 NASA	(I)	(II)	(III)
SO2 HyADS Actual	3.594***	1.283***	0.635***
	(0.189)	(0.150)	(0.096)
Constant	3.071***	3.437***	6.589***
	(0.052)	(0.034)	(0.202)
Year FE	No	Yes	Yes
Incinerator FE	No	No	Yes
R^2	0.447	0.817	0.959
N	451	451	451

Table A.2: SO2 Concentration from Monitors as a Function of SO2 Concentration from the NASA. This table shows the results for OLS specifications in which the dependent variable is the annual average concentration of SO2 (in μg/m^3) measured by the monitors of the University of Aarhus. The main explanatory variable is the annual average concentration SO2 (in μg/m^3) from the MERRA-2 dataset provided by the NASA. Each monitor is matched with a parish depending on its coordinates; the grid of the NASA (0.5^{o}x0.625^o) is matched with the shapefile of the Danish parishes. We control for year and monitor fixed effects. Standard errors are in parentheses; *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

SO2 Monitors	(I)	(II)	(III)
SO2 NASA	2.136***	1.891***	2.623***
	(0.099)	(0.201)	(0.191)
Constant	-6.696***	-5.033***	-6.147***
	(0.751)	(1.387)	(2.544)
Year FE	No	Yes	Yes
Monitor FE	No	No	Yes
R^2	0.814	0.937	0.983
N	107	107	107

A.2 Alternative Versions of HyADS Based on Other Pollutants

In addition to considering HyADS based on SO2 emissions as our main measure of exposure to incinerators, we also compute HyADS based on emissions of PM2.5; in the future, we plan to compute other versions based on emissions of dioxins and heavy metals. We attach a lower degree of reliability to these versions of HyADS, given the uncertainty surrounding our data on the emissions of these pollutants.

As for PM2.5, we never observe it on the Green Accounts, which only contain data on PM10. In addition, concentrations of PM2.5 from the NASA are estimated with a higher degree of uncertainty; while in the case of SO2 the NASA directly reports concentrations, in the case of PM2.5 the NASA only estimates the concentrations of the precursors, and the concentrations of the pollutant of interest need to be computed by applying a formula specified on the website of the MERRA-2 dataset[40]. As a result, according to Table A.3, PM2.5 from the NASA performs rather poorly as a predictor of PM2.5 from monitors (the R^2 for the specification that only controls for the NASA concentration is just 0.2). As for dioxins heavy metals, we do not have any data on concentrations either from the NASA or from the monitors, which leaves us with imputed emissions as our only alternative to the Green Accounts.

[40]PM2.5 = DUSMASS25 + OCSMASS+ BCSMASS + SSSMASS25 + SO4SMASS* (132.14/96.06); see https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/FAQ/

Table A.3: PM2.5 Concentration from Monitors as a Function of PM2.5 from the NASA. This table shows the results for OLS specifications in which the dependent variable is the annual average concentration of PM2.5 (in μg/m^3) measured by the monitors of the University of Aarhus. The main explanatory variable is the annual average concentration of PM2.5 (in μg/m^3) from the MERRA-2 dataset provided by the NASA. Each monitor is matched with a parish depending on its coordinates; the grid of the NASA (0.5^{o}x0.625^o) is matched with the shapefile of the Danish parishes. We control for year and monitor fixed effects. Standard errors are in parentheses; *, **, and *** mean statistically significant at the 10, 5, and 1% levels.

PM2.5 Monitors	(I)	(II)	(III)
PM2.5 NASA	1.616***	1.139*	-2.266
	(0.451)	(0.625)	(1.582)
Constant	-1.476	3.110	41.945**
	(4.356)	(6.011)	(15.763)
Year FE	No	Yes	Yes
Monitor FE	No	No	Yes
R^2	0.204	0.538	0.904
N	52	52	52

A.3 Before-and-After Comparison in the Sample of Closings

While the specification fitted to the sample of openings relies on a difference-in-differences design, in the sample of closings we observe individuals since the closing year, without performing any before-and-after comparisons. Here, for robustness, we extend our analysis to a framework amenable to a before-and-after comparison.

More in detail, we build the sample of closings by following the same rationale as before; however, we include individuals who live in the relevant municipalities and die before the closing year (actual or fictitious); we observe both exposed and unexposed individuals since birth/1982 (not since the closing year) up to death/2018.

As is the case in the main text, the main variables are three: cumulative exposure, intensity, and latency. Cumulative exposure is the cumulative sum of HyADS, computed over an individual's lifetime; given that the first incinerator included in our dataset was activated in 1950, the sum starts from the most recent between birth and 1950. This variable is measured only with reference to the exposed and is divided by 0.9, being expressed in terms of 0.9-HyADS-Years. For person-years observed before the closing year, cumulative exposure is time-varying; after the closing year, it is fixed and equal to the amount observed in correspondence to the closing year. The intensity of exposure is measured as cumulative HyADS divided by current age or the difference between the current year and 1950 for person-years observed before the closing year and as cumulative HyADS at closing divided by age at closing or the difference between the closing year and 1950 for person-years observed after the closing; 1950 is used in place of the birth year for individuals born before 1950. Finally, latency is measured in terms of years since closing.

In Figure A.2, we plot the average HyADS for exposed individuals (divided into three categories of average intensity, using the usual cutoffs of 0.4 and 0.9 HyADS units) and unexposed individuals as a function of time since closing. The event study is well-designed, given that HyADS is significantly higher for exposed individuals before the closing, when it drops to a level indistinguishable from the one associated with the unexposed.

We fit a Cox regression that allows for a before-and-after comparison. The failure event is represented by lung cancer. The main explanatory variable is cumulative exposure; this

variable is continuous and its marginal effect on lung cancer is let vary with the categories of two variables, years since closing ($<$0, 0-4, 5-9, 10-14, and 15+ years) and average intensity ($<$0.4, 0.4-0.9, and $\geq$0.9 HyADS units). As usual, we control for income and wealth quintiles lagged by two years as well as for marital status. We consider age as the time scale and we stratify the sample by sex, birth cohort, and pair of municipalities.

Table A.4 documents a pattern very similar to the one identified in the main text. Namely, we find evidence that latency is shortened by the intensity of exposure. Indeed, only for the highest category of intensity we document that cumulative exposure increases the probability of getting lung cancer even before the closing; for the lowest category of intensity the effect of cumulative exposure on lung cancer materializes between 10 and 15 years since the closing; for the second, between 5 and 15 years. For all categories of intensity, as we find in the main text, the effect becomes statistically insignificant 15 years after the closing.

Figure A.2: Average HyADS for the Exposed (Divided by Category of Intensity) and the Unexposed in the Sample of Closings. This picture shows the average number of HyADS units as a function of years since closing, distinguishing between exposed individuals (divided into categories of average intensity of exposure, with the cutoffs being 0.4 and 0.9 HyADS units) and unexposed individuals. The sample is built by considering the closings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was shut down, before or during the closing year. Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious closing year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime.

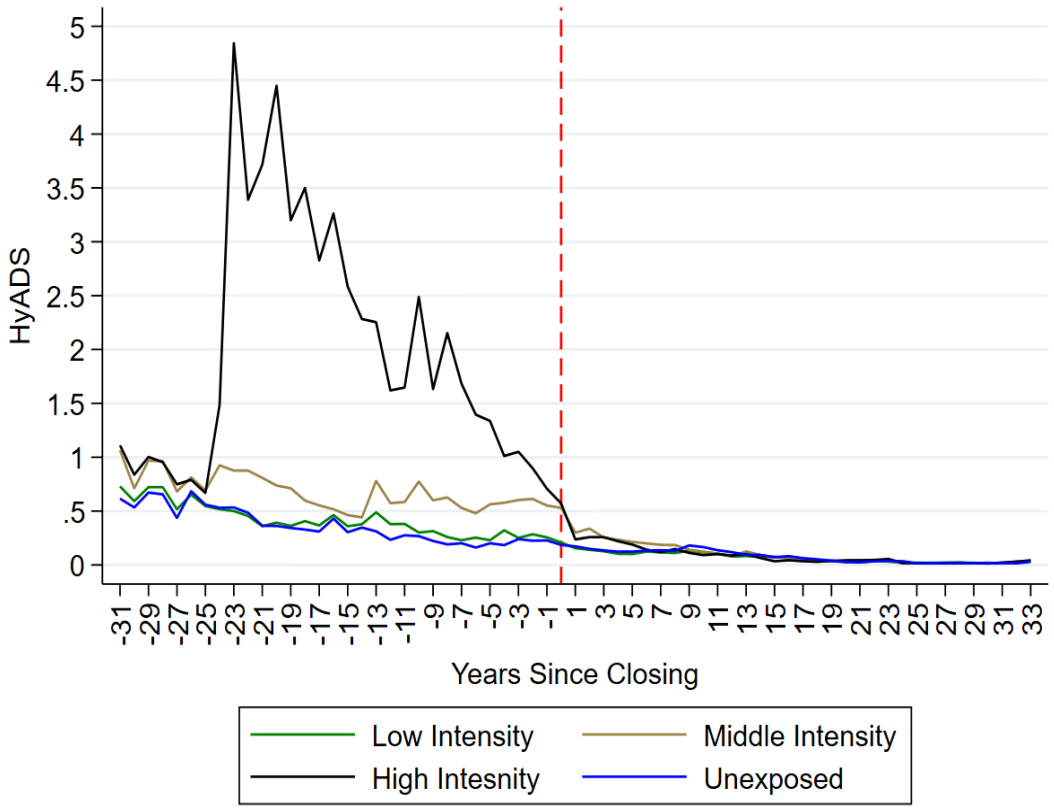


Table A.4: Hazard Ratio for 1 More 0.9-HyADS-Year as a Function of Time Since Closing and Average Intensity of Exposure Before the Closing in a Before-and-After Comparison. This table shows the results for a Cox regression in which the dependent variable is represented by lung cancer and the main explanatory variable is the amount of HyADS accumulated by an exposed individual from the most recent between birth and 1950; for person-years observed before the closing, the cumulative sum goes up to the year of interest; for person-years observed after the closing, the sum goes up to the closing year; the sum is divided by 0.9 and measured in 0.9-HyADS-Years. The main explanatory variable is interacted with categories of years since closing and with categories of the average number of HyADS units per year; the latter is computed by dividing cumulative HyADS by the minimum of current age or age at closing and the difference between the current year or the closing year and 1950, depending on whether a person-year is observed before or after the closing and on whether an individual was born before or after 1950. The sample is built by considering the closings of incinerators between 1982 and 2018. We restrict the attention to the individuals who never change parish during their lifetime. The exposed are identified as individuals living in a municipality where an incinerator was shut down, before or during the closing year. Each municipality involved in a closing is matched with a municipality where an incinerator has never operated, by population and area. The unexposed are identified as individuals living in an unexposed municipality, before or during the fictitious closing year that depends on the matched municipality. We drop the exposed who are associated at least once with a level of HyADS of at least 0.9 units after the closing year; we drop the unexposed who are associated at least once with a level of HyADS of at least 0.9 units during their entire lifetime. We consider both the exposed and the unexposed as under observation since birth/start of the sample of period. We control for household income and wealth quintiles lagged by two years as well as for marital status; we consider age as the time scale; we stratify the sample by sex, birth cohort, and pair of municipalities. The table shows the hazard ratios for the variables of interest; each hazard ratio must be interpreted as the gross excess risk associated with 1 more 0.9-HyADS-Year of cumulative exposure, accounting for the number of years elapsed since the closing and for the average number of HyADS units per year; standard errors are in parentheses; *, **, and *** mean statistically different from 1 at the 10, 5, and 1% levels. The data on median 0.9-HyADS-Years are from Table 12.

Average Intensity	Years Since Closing	HR Per 0.9-HyADS-Year	Median 0.9-HyADS-Years
0-0.4 HyADS	Before Closing	0.9987	
		(0.0023)	
	0-4 Years	1.0001	
		(0.0035)	
	5-9 Years	1.0037	8.6625
		(0.0043)	
	10-14 Years	1.0144**	
		(0.0065)	
	15+ Years	1.0025	
		(0.0056)	
0.4-0.9 HyADS	Before Closing	1.0024	
		(0.0018)	
	0-4 Years	1.0021	
		(0.0034)	
	5-9 Years	1.0074*	24.7128
		(0.0039)	
	10-14 Years	1.0109**	
		(0.0047)	
	15+ Years	1.0033	
		(0.0042)	
$\geq$0.9 HyADS	Before Closing	1.0115***	
		(0.0011)	
	0-4 Years	1.0096***	
		(0.0021)	
	5-9 Years	1.0069***	81.0938
		(0.0025)	
	10-14 Years	1.0039	
		(0.0032)	
	15+ Years	1.0018	
		(0.0044)	
	N	7,293,033	

A.4 Back-of-the-Envelope Calculations

A.4.1 Sample of Closings

Our first aim is to translate the hazard ratios presented in Table 12 into excess lung cancer cases within the sample of closings, distinguishing among categories of average intensity and categories of time since closing.

To compute excess lung cancer cases, we need to have an estimate of the excess risk of lung cancer caused by incinerators. The first step is to decide on the baseline risk of lung cancer. Then, given that Table 12 presents the hazard ratios associated with 1 more 0.9-HyADS-Year accumulated before the closing, the second step is to have an estimate of the overall amount of cumulative exposure accumulated by the exposed individuals during the pre-closing period. Given these two estimates, considering the multiplicative interpretation of the Cox regression, excess risk can be estimated as,

$$\text{ExcessRisk} = \text{BaselineRisk} \cdot (\text{HazardRatio}^{\text{CumulativeExposure}} - 1)$$

As for the first step, we opt for 20 basis points, which is a conservative approximation of the sample average of lung cancer in the sample of closings (22.58 basis points). As for the amount of exposure accumulated before the closing, we go for the median value of this variable associated with the individuals included in each category of intensity; we do not consider the mean given that the distribution of cumulative exposure can be highly right-skewed due to the presence of particularly polluting plants in the sample.

For the first category of intensity (0-0.4 HyADS units), the median cumulative exposure at the time of closing is 8.66 0.9-HyADS-Years. This implies that for these individuals the annual excess risk of lung cancer is $20 \cdot (1.0144^{8.66} - 1) = 2.64$ basis points between 10 and 15 years since the closing. Following the same rationale, for the second category, we have $20 \cdot (1.0068^{24.71} - 1) = 3.65$ basis points every year between 5 and 10 years since the closing, and $20 \cdot (1.0106^{24.71} - 1) = 5.95$ basis points between 10 and 15 years since the closing. Finally, for the highest category of intensity, we have $20 \cdot (1.0074^{88.09} - 1) = 18.29$ basis points between 0 and 5 years since closing and $20 \cdot (1.0057^{88.09} - 1) = 13.00$ basis points between 5 and 10

years.

Given our estimates of the excess risk, the overall number of excess cases can be found by considering the size of the relevant populations. In total, in the sample of closings, there are 149,539 unique exposed individuals, with 116,490, 24,086, and 8,963 individuals belonging to the different categories of intensity. Based on the above excess risks of lung cancer computed in basis points, we estimate the overall number of excess lung cancer cases observed after the closing and due to exposure to pollution from incinerators before the closing to be equal to 410. Namely, for the first category of intensity, the total number of excess lung cancer cases is $116,490 \cdot 2.64 \cdot 5/10,000 = 154$, where 5 is the number of years for which we estimate the excess risk of lung cancer to materialize (between 10 and 15 years since the closing in this case); for the second category, we have $24,086 \cdot 3.65 \cdot 5/10,000 + 24,086 \cdot 5.95 \cdot 5/10,000 = 44 + 72 = 116$; for the third, we have $8,963 \cdot 18.29 \cdot 5/10,000 + 8,963 \cdot 13 \cdot 5/10,000 = 82 + 58 = 140$. Note that only 82 excess lung cancer cases are estimated to materialize within 5 years after the closing. Assuming that these 82 cases are uniformly distributed within the 5 years, we have that only 16 out of the total 410 excess lung cancer cases occur at the time of the closing. In consequence, not accounting for latency (i.e., limiting the analysis to the closing year, under the assumption that the effect of cumulative exposure vanishes after people stop being exposed) would lead to capture only 4% of the overall number of excess lung cancer cases (16/410) related to the post-closing period but caused by exposure accumulated before the closing. The magnitude of the underestimation depends on the intensity of exposure. For the two lowest categories of intensity, 100% of the excess lung cancer cases remain undetected if latency is not accounted for. For the third category, around 90%.

A.4.2 Assessing the Benefits of Regulation

In addition to quantifying the impact of incinerators on health in the sample of closings, Table 12 can be used to perform back-of-the-envelope calculations related to our entire dataset covering the universe of Danish people observed between 1982 and 2018. Our aim is to study the health effect effects of incinerators over time. Namely, we want to evaluate the benefits of the two emissions caps introduced in 1991 and in 2003 (see Subsection 2.2).

In 1990, before both emissions caps, the incinerator associated with the median value of HyADS per year (0.8) was in Svendborg. This municipality is particularly suitable for our purpose because it has a representative number of residents (around 50,000), given the Danish population and the number of Municipalities (5 million people and 98). As in 1990 the plant of interest fell within our second category of intensity (0.4-0.9 HyADS units) and the median resident observed in 1990 in the relevant municipality had an amount of cumulative exposure of 22.90 HyADS units, considering a baseline lung cancer risk of 20 basis points for consistency with the back-of-envelope calculations performed before, we have that in one more year of activity the plant was generating $20 \cdot (1.0068^{(22.9+0.8)/0.9} - 1) - 20 \cdot (1.0068^{22.9/0.9} - 1) = 0.15$ basis points of excess risk with a latency of 5-10 years (equivalent to 3.75 excess cases in the population of 50,000 residents) and $20 \cdot (1.0106^{(22.9+0.8)/0.9} - 1) - 20 \cdot (1.0068^{22.90/0.9} - 1) = 0.25$ basis points of excess risk (6.25 excess cases) within 10-15 years. Then, considering 50% of the population of 50,000 residents (which is roughly the attrition rate that we experience in the sample of closings used to estimate our hazard ratios)[41], we obtain that the overall number of excess cases is $0.5 \cdot (3.75 + 6.25) = 5$ with a latency of 5-15 years.

In 2004, right after the second emissions cap, the Svendborg incinerator produced 0.09 HyADS units (slightly below the median of 0.15). Following the same rationale as above, considering that the incinerator started to belong to our first category of intensity (0-0.4 HyADS units), and that the median cumulative exposure for the residents was 22.85, we obtain 0.5 excess lung cancer cases with a latency of 10-15 years.

Considering 10 million dollars as the value of a statistical life and assuming that 80% of the excess lung cancer result in excess deaths[42], we have that the median incinerator observed at the beginning of our sample was determining a human cost of 40 million dollars per year; in 2004, this cost dropped to 4 million dollars. According to the financial accounts of the Svendborg incinerator[43], the plant generated revenues of around 10 million dollars in 2021; more broadly,

[41] Note that the average of the lung cancer dummy over our entire dataset covering all Danish people is 10 basis points, half the average obtained with reference to the sample of closings; this is due to the fact that in the sample of closings we do not consider the people born after the event of interest; accordingly, in this context, dividing the number of excess cases by 2 is also equivalent to computing this number in accordance with the overall baseline risk of lung cancer.

[42] https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html

[43] https://datacvr.virk.dk/enhed/virksomhed/33054521?fritekst=33054521&sideIndex=0&size=10

the entire waste incineration industry generates an average of 900 million dollars of revenues per year in Denmark[44], with the yearly number of plants being roughly 30 (now 24). Accordingly, we can safely state that the median incinerator observed in the early part of our sample every year determined an overall externality in terms of excess deaths at least as large as the amount of revenues currently observed (we unfortunately do not have access to historical financial accounts). In 2004, after the two emissions caps, this externality accounted for 40% of the revenues.

We conclude that even if the median incinerator has reduced its impact on human health by a factor of 10 (from 5 to 0.5 excess lung cancer cases per year of activity) over time, the long-term impact of its environmental externalities still accounts for around 40% of its revenues.

[44] https://www.statista.com/forecasts/896944/waste-treatment-and-disposal-revenue-in-denmark