LUND UNIVERSITY
School of Economics and Management

Master program in Economic Demography

The Effect of Fine Particle Pollution on Mortality and Life Expectancy in Sweden 1990-2011

Lisa Carlson
lisa.carlson.003@student.lu.se

Abstract: The level of fine particle pollution in the air has been found to be damaging to health. This study aims to see if there is a negative correlation between the level particulate matter and life expectancy, and a contrary correlation between particulate matter and cause-specific mortality, using a statistical model with the 21 Swedish counties. The outcomes from the statistical models do not, however, show much indication that the pollution level has a significant effect on either life expectancy or mortality. If further research would be pursued on fine particle pollution on life expectancy and mortality in Sweden, the use of smaller geographical areas and a longer time frame could improve the analysis and might lead to more conclusive results.

Key words: Air pollution, Mortality, Sweden

EKHM51
Master thesis (15 credits ECTS)
June 2015
Supervisor: Martin Dribe
Examiner: Jonas Helgertz
Word Count: 15855

Table of Contents

Introduction:

Human health and mortality are affected by a vast number of factors, many of which are man-made. Emissions of a range of different pollutants, for example, have a negative impact on life expectancy. The interest in the effect of pollution on the environment and human health is currently in vogue and reports of the effect and research on the subject are plentiful. The analysis in this essay focuses on particulate matter, a fine particle air pollutant that is hazardous to health and over long term exposure can be the cause of respiratory and/or cardiopulmonary diseases and lead to premature death. The effect of very fine particle pollutants, called PM_{25} because of their size, on life expectancy and mortality is the focus of the analysis in this essay. These very fine particles are more hazardous to health than the fine particles (PM_{10}). Previous research on the correlation between fine particle pollution and life expectancy in the United States and in the European Union show that the level of pollution in the air is negatively correlated with the longevity of life.

The aim of this analysis is to see if a similar relationship exists between PM_{25} and life expectancy in Sweden. The article in previous research that was the main inspiration for the method of analysis in this essay was an article by Correia et. al. (2013) that investigated the correlation between PM_{25} concentrations in the air and life expectancy in more than 500 US counties. Since there are many ways in which the United States and Sweden are different, the hypothesis for the correlations between the air pollutant and life expectancy differs from the results of the Correia et. al. (2013) study. There are also certain limitations in this analysis that were not encountered in that study. Lack of data for Sweden is an example of a limitation that may affect both the hypothesis and the findings of this analysis. The lack in data can be either years available or the form the data is available in.

The aim of this essay is to answer the following research questions: does the PM_{25} pollution affect life expectancy negatively in Sweden? And if not, what conditions or factors exist in Sweden that may lead to a different relationship when compared to the United States?

In contrast to the previous research, a potential positive correlation between fine particle pollution and cause-specific mortality is also considered. Cause-specific mortality (deaths due to respiratory and/or cardiopulmonary diseases) is a mediator between PM_{25} and life expectancy. If PM_{25} is expected to have a negative effect on life expectancy, it is expected to have a positive effect on cause-specific mortality.

The essay is divided as follows: background to the problem and history of the development of fine particle pollution and life expectancy; previous research and theory; method with definition of data, descriptive statistics of data and outline of statistical model; results from the statistical analysis; discussion of the result in terms of theory and previous research; conclusion and further research.

Background:

Fine particle pollution, referred to as $PM_{2.5}$ because of their size (≤2.5 μm in diameter), is hazardous to both the health of the human population and to the environment (epa.gov, 2015). The fine particles are chemical reactions of different gases in the air, such as sulfur dioxide and nitrogen dioxide (Ibid.). The emissions are widespread and come from a multitude of different sources such as vehicles and various kinds of production (epa.gov, 2014). Power plants, gasoline and diesel engines, wood combustion and high temperature industrial processes (such as steel mills) are examples of man-made emissions but the pollutants can also be emitted by natural sources during forest fires (which may be man-made or natural)(Ibid.). Because of the small size of the particles, sources in one part of the country are able to influence the environment hundreds of miles away (Ibid). In the South of Sweden, for example, pollution from continental Europe is spread by e.g. the transportation of goods.

Fine particle emissions affect the health of the population by their ability to penetrate deep into the human lung and are therefore the cause of respiratory illnesses such as asthma (Ibid.). The particles are also the cause of premature death due to these respiratory illnesses (Ibid.). Preexisting conditions such as cardiopulmonary disease increase the risk of death due to $PM_{2.5}$ emissions, especially for the already more morbidity susceptible elderly (Ibid.). Children are also more at risk due to not having fully developed respiratory systems and because children breathe in more air than adults, proportionally (Ibid.). Smokers are also more l susceptible to the pollution. Correlations between air pollution and cardiopulmonary diseases have also been found (Meister et. al., 2012).

Because of the adverse effects of $PM_{2.5}$ emissions organizations such as the World Health Organization has implemented goals for emission regulation (who.gov, 2014). The WHO Air Quality Guidelines states that the yearly average of $PM_{2.5}$ emissions per 24 hours should be less or equal to 10 $\mu g/m^3$ (Ibid.). Individual nations may have different target levels. In this essay Swedish emission data is

regressed on life expectancy to see how the level of pollution affects mortality. The analysis in this essay is based on several articles on the same subject, but mainly a study of the $PM_{2.5}$ emission levels in hundreds of US cities by Correia et. al. (2013). The regulation targets in these two nations differ: in Sweden the target pollution level is the same as in the WHO Air Quality Guidelines, 10 $\mu g/m^3$ (miljomal.se, 2014). In the US, on the other hand, the National Ambient Air Quality Standards state that for the primary standard (which is the standard for human health protection) the target is 12 $\mu g/m^3$ (epa.gov, 2014, a).

History of $PM_{2.5}$ Emissions:

Recorded levels of $PM_{2.5}$ in the air in Sweden is scarce, but the recorded figures that are available show that the levels are too high in larger cities and in large parts of the South of Sweden (miljomal.se, 2014). In the northern part and in rural areas the levels are within the target of 10 $\mu g/m^3$ for most of the years since year 2000 (Ibid.). Traffic, industry and population density are also higher in the South of Sweden. For the US, the national average has been below the target of 12 $\mu g/m^3$ for the last years (epa.gov, 2014, a). The EPA has readily available data for large parts of the United States which makes the analysis of the effect of $PM_{2.5}$ levels on mortality possible using the $\mu g/m^3$ averages (Ibid.). Because of the scarcity of recorded levels in this measure for Sweden, however, the history of $PM_{2.5}$ emissions is shown using $PM_{2.5}$ emissions in thousands of tons, data that is available for the nations as a whole (as seen below using information gathered from OECD) and for the 21 individual Swedish counties (as used in the statistical analysis that follows, gathered from Naturvårdsverket). The concentration of pollution in the air is therefore disregarded and the amount of pollution emitted is used to show the development of the pollution level since 1990. Lower levels of emitted particles should, in theory, lead to a lower concentration in the air but other factors may affect this.

The following graphs show the difference in *emitted* fine particles in Sweden and in the United States. The graphs are configured in such a way that the

development of emissions is more noticeable. This is managed by using two axes with different scales.

Figure 1: PM_{25} Emissions in Sweden and the United States 1990-2011:

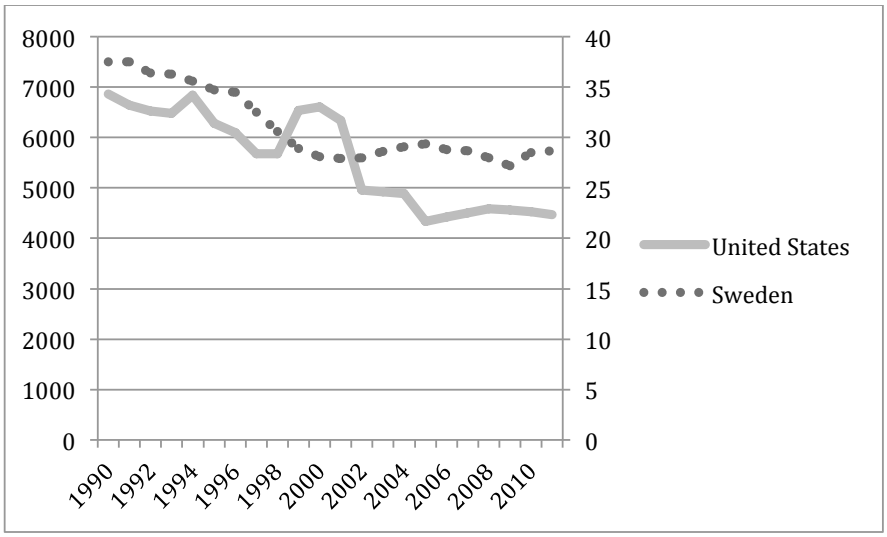

Figure 1: PM2.5 Emissions, Sweden and the United States, Source: stats.oecd.org

The figure above, Figure 1, depicts the fine particle, $PM_{2.5}$, emissions in Sweden and in the United States over the time period 1990-2011. The emissions are in thousands of tons and the diagram has two axes (Sweden on the right and the US on the left) to show the development throughout the years more easily since the size differences of the countries are vast.

To view the fine particle emissions for the two countries without (some of) the large differences between the nations, emissions in thousands of tons per capita can be used. The figure below, figure 2, shows the development of the $PM_{2.5}$ emissions per capita for the years 1990-2011.

Figure 2: PM_{25} Emissions per capita in Sweden and the United States 1990-2011

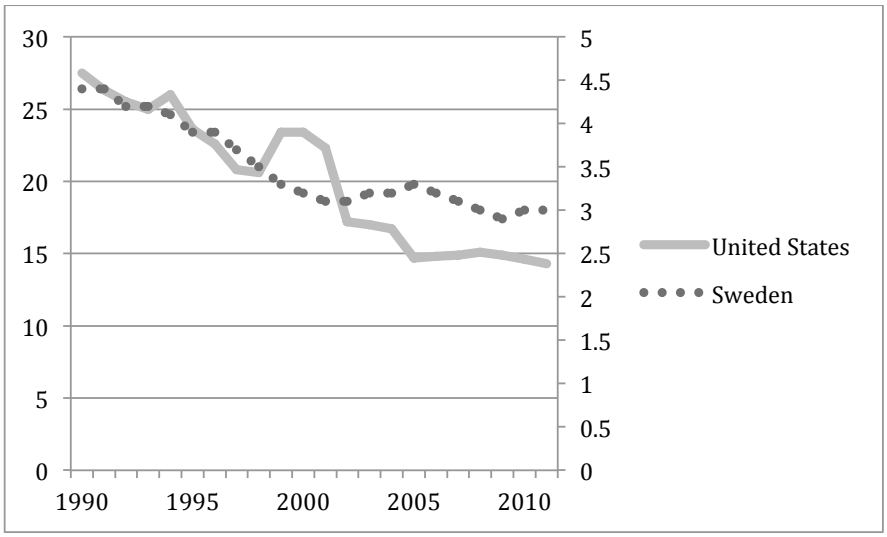

Figure 2: Emissions per capita, Sweden and the United States, Source: stats.oecd.org

The data is still divided on two axes (Sweden on the right, the US on the left, as in figure 1) since the development in the particle emissions is still easier to compare in this manner.

The emissions in the US are still several times higher than the emissions in Sweden.

A second way to see the difference in emissions between the counties is to look at the emissions in kilograms per USD 1000. The unit of comparison is emissions per unit of gross domestic product in kilograms per USD 1000.

Figure 3: PM_{25} Emissions per USD 1000 in Sweden and the United States 1990-2011

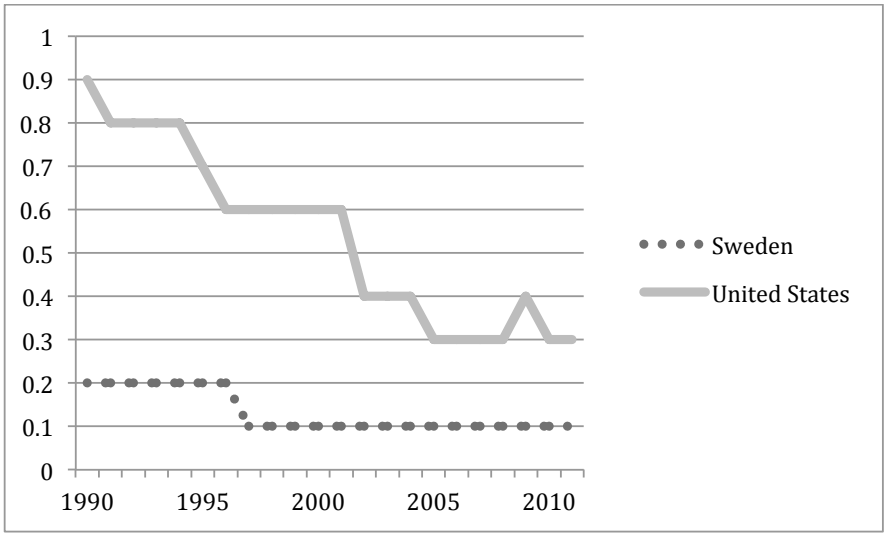

Figure 3: Emissions in kg per USD 1000, Sweden and the United States, Source: stats.oecd.org

In figure 3, the two series share one y-axis, in contrast to the previous two figures.

Regardless of in what way the emission development through the years is depicted in the three diagrams above, the emission levels are higher in the United States than in Sweden. This is important for the basis of the theory since a lower level of pollution to begin with should affect health and environment less than a higher level would. Any results obtained regarding the effect that fine particle pollution has on mortality is therefore expected to be lower for Swedish data than for American data.

History of Emission Regulation:

The European Union began to regulate the emissions of $PM_{2.5}$ and PM_{10} in the beginning of the 1980s (Wolff & Perry, 2010). The limits in the 1980s were 80 $\mu g/m^3$ as a daily mean in the summer months and 130 $\mu g/m^3$ in the winter

months (Ibid.). In 2005 the limit was lowered to 50 µg/m^3 on no more than 35 days yearly and an annual average of less than 40 µg/m^3 (Ibid.). The new regulations stipulated that an even lower limit would be implemented in 2010 of 20 µg/m^3 as a yearly average and only 7 days annually with 50 µg/m^3 (Ibid.).

History of Life Expectancy (both sexes):

Figure 4: Life Expectancy in Sweden and the United States 1990-2011

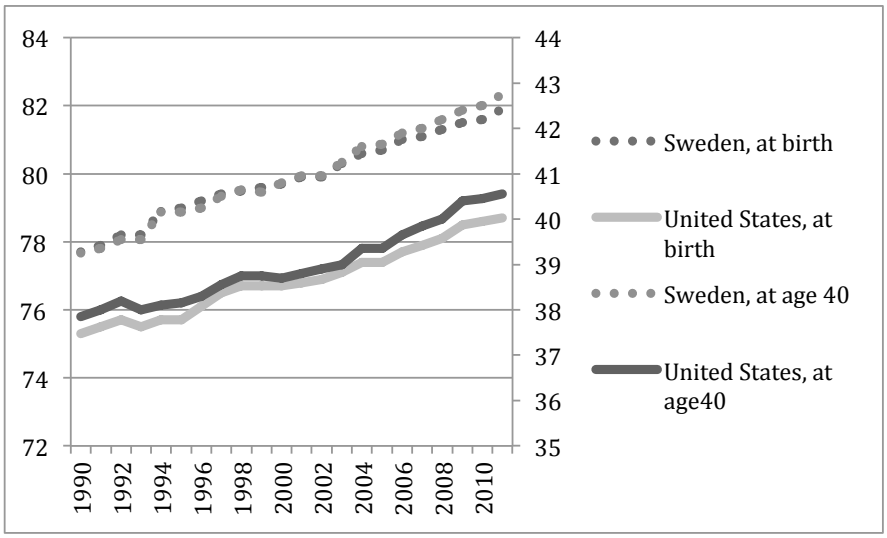

Figure 4: Life Expectancy, Sweden and the United States, Source: stats.oecd.org

Life expectancy has risen in both the United States and in Sweden during the years in this analysis. Figure 4, above, shows the increase in life expectancy at birth and life expectancy at age 40 during the years 1990 to 2011. Life expectancy at birth is depicted on the left and life expectancy at 40 is depicted on the right to show the developments in both more easily in the same graph. There are positive developments for both countries but the starting point for Sweden is higher for both life expectancy at birth and at age 40. This discrepancy can, just as the lower emissions level, influence how much a decrease in $PM_{2.5}$ emissions affects mortality.

If life expectancy is already high, a decrease in emissions is likely to have less of an effect on life expectancy. The same is true for cause-specific mortality.

Theory:

The hypothesis is that a reduction of air pollutants in the form of fine particles, $PM_{2.5}$, in the air in different Swedish counties should lead to decreases in the mortality of its residents or put differently, decreases in pollution level should lead to increases in life expectancy. Several studies of areas of the United States have shown that lower levels of air pollution are correlated with higher levels of life expectancy (Harvard Six Cities Study; Medicare Study; American Cancer Society Study)(Dockery et. al., 1993; Pope et. al., 1995).

Previous research (see below) has shown that exposure to fine particle pollution, and particularly very fine particle pollution, is harmful to health. The pollution is mostly hazardous to the respiratory system but may also bet the cause of cardiopulmonary diseases. It can also aggravate those diseases if preexisting. The level of pollution is caused by several factors such as industrial factors (production and transportation of goods) and human factors (urbanization and use of cars). In Sweden, and in other parts of the world, the use of studded tires, road salt and traction sand during the winter months increases the level of fine particle pollution in the air (Meister et. al., 2012).

The regulation of production may help reduce the level of particle pollution in the air. Since higher levels of production lead to higher levels of emissions, an area with more industry and a more prosperous industry may also have higher levels of fine particle emissions. This is given that the level of emissions is higher than zero, which is to be expected. The state of the economy may thus affect the production level and thereby the emission level. A “good” economy may thus be worse for the environment and health than a “bad” environment. The business cycle might level out the effect of a “good” and a “bad” economy.

A higher level of production may also lead to higher transportation of produced goods. The choice of transportation may thus affect the level of emission differently. The transportation of goods also helps transport air pollutants from one side of a country to another or between countries. The southern part of Sweden see more traffic than the northern part of Sweden and therefore the

pollution level in the south should, in theory, be higher than the level in the north. Skåne and Västra Götaland, especially, have higher levels of transportation of goods and Västra Götaland, and in particular, Gothenburg, have the highest levels in the country (according to figures from 2000)(scb.se, 2015).

Industrialization and deindustrialization affect the level of emissions in the air. Urban areas have higher levels of pollution than rural areas. Urban areas see more production than rural areas, usually, and the level of traffic is higher in more densely populated areas. The use of alternative transportation to cars may lessen the effect of a higher population but the kind of public transportation matters. In more rural areas the availability of alternative transportation to cars may also be scarcer. Industrialization may lead to population increases and depending on the kind of production in the area the education level of the population may change. In Sweden (and probably elsewhere) education levels are higher in more populous areas. Deindustrialization may, contrarily, lead to population decreases. Population change, in either direction, may affect the emission levels in the area.

Though urban areas may see more air pollution than rural areas, how affected their respective population is differs. The young and the elderly, for example, are more susceptible to health problems so an area where the average age is high may also have a population where a higher proportion suffer and die from diseases affected by air pollutions. Smokers are also more affected by emissions than non-smokers. In Sweden, people in a rural area consume more tobacco products than a person in an urban area (scb.se, 2015). For 2000, people in smaller cities and less populated areas spent more of their income on tobacco products than people n larger cities (Ibid.). People in smaller towns in the North of Sweden, in particular, consumed more tobacco products than anywhere else in the country (Ibid.). The tobacco products are not limited to cigarettes, however. The mortality from health problems due to smoking and from health problems caused by fine particle pollution may be higher in an area with a lower level of pollution but with a higher percentage of smokers. This can affect life

expectancy and a county with low life expectancy does not have to be a high pollution county. There are, naturally, more factors in play.

The Correia et. al. (2013) article that analyzes the effect of air pollution control on life expectancy in the United States from 2000 to 2007 is used as inspiration for the theory. That article, however, tried to see if the air pollution regulations for those years are less effective than the regulations implemented in the 1980s (Correia et. al., 2013). The analysis in this essay uses data from the year 1990 until today and does not aim to compare different levels of regulation. The level of regulation has, on the other hand, changed in Sweden since 1990 and though not looked at from the perspective of the effectiveness of regulation, the more intense regulation may lead to, and should lead to, lower levels of emissions which could cause a substantially smaller effect of pollution on life expectancy or cause-specific mortality.

In addition to analyzing the effect that PM_{25} pollution has on life expectancy, the analysis in this essay focuses on the effect that air pollution has on cause-specific mortality. Deaths due to respiratory illnesses or cardiopulmonary diseases represent the cause-specific mortality and are referred to as "cause of death" in the statistical analysis. The theory is that cause of death is positively affected by fine particle pollution and that cause of death then is negatively correlated with life expectancy. It should, hypothetically, act as an intermediary between pollution levels and life expectancy.

The United States and Sweden do have similar characteristics. Some aspects of their populations and their economies are similar: they have high life expectancies and they are both advanced economies. There are however differences between them that might be important to consider when using a method similar to that of the Correia (2013) article. The population of Sweden might be considered more homogenous than the population of the United States so the use of the proportion of white/black/Hispanic would not tell much in the case of Sweden.

The pollution level of the two countries differs as well as aforementioned (see Figure 1-3); the air pollution emissions, as measured in thousands of tons, are lower in Sweden than in the United States. The levels as measured in $\mu g/m^3$ are on average lower in the United States than in Sweden, as of recent years. So, the concentration in the air is higher in Sweden (in urban areas but not in rural areas) but emission levels are higher in the US.

The pollution level and the level of life expectancy vary in the Swedish counties. The counties with higher population density, education level and income are also counties where the level of $PM_{2.5}$ is higher than the Swedish average. There are also counties where the opposite is true: some Swedish counties have low levels of fine particle emissions yet have lower than average life expectancy. These differences may affect the result of the statistical models since the higher level of pollution may seem to be correlated with a higher level of life expectancy though the opposite should logically, and barring all other factors, be true.

In sum, the hypothesis is that higher levels of fine particle emissions lead to lower life expectancy. The theory for cause-specific mortality, contrarily, is that $PM_{2.5}$ emissions have a positive effect on mortality since higher emissions lead to higher numbers of deaths. Other factors than air pollution affect life expectancy and thus fine particle emission is not the sole causation of life expectancy.

Overview of Previous Research

There is a lot of research on the affects of pollution on health and many studies that show that the quality of the environment affects morbidity and mortality (Mariani et. al., 2010). There are also several studies of air pollution as a public health burden (Pascal et. al. 2013). The research do show that the level of pollutants in the air has been decreasing since the 1950s (at least in the European countries), but that the existing level still does have severe negative effects on the health of the population (Ibid.).

There are several studies about the effects of air pollution on morbidity and mortality in the United States. The Harvard 6 Cities Study and The American

Cancer Society Study sought to see if fine particle air pollution ($PM_{2.5}$) had an adverse effect on health and found that the risk of cardiopulmonary mortality did increase with the increase of fine particle pollutants in the air (Correia et. al. 2013). The effect of air pollution is more potent in the long run than in the short run, however; a study of two cohorts in Scotland showed that mortality did increase because of air pollution but that the effect was larger in the long run (Beverland, et. al. 2012)

There are research on air pollution and mortality and/or life expectancy in the United States that use the data made available by the EPA and other sources, that point to the conclusion that reductions in air pollution affect mortality and life expectancy positively (Samet, 2013; Correia et. al., 2013; Mariani et. al., 2010; Liu, 1979). Correia et. al. (2013), for example, used cross-sectional data on the emissions of $PM_{2.5}$ and life expectancy in 545 counties in the US and using a regression model, controlling for factors such as smoking, that lower levels of $PM_{2.5}$ lead to longer life expectancy. They found that the life expectancy, on average, increased by 0.84 years when $PM_{2.5}$ decreased by 1.56$\mu g/m^3$ from year 2000 to year 2007. Their model also included factors such as the mortality due to COPD. Samet (2013) continued to build on the results found by Correia et. al. (2013), and found that the gains in life expectancy were greater in urban areas. Samet (2013) also suggests that further gains in life expectancy are possible with further decreases in the emission levels of $PM_{2.5}$.

A study done by Chay et. al. (2003), on the changes in air pollution due to the regulations of the Clean Air Act in 1971, compared to 1970, indicated that the regulations did not affect mortality or health in the short run. The conclusion that the regulations of air pollution are more effective in the long run is in line with the conclusion of the study of two cohorts in Scotland (Beverland, et. al. 2012).

In Europe, the Asphekom project has studied the urban air pollution and its effect on the population in 25 European cities in countries in the European Union (Pascal et. al. 2013). They compared the levels of different air pollutants ($PM_{2.5}$,

PM_{10}, Ozone) in the chosen 25 cities to the World Health Organization's Air Quality Guidelines (WHO-AQG) to see if the cities complied to those guidelines (Ibid.). For $PM_{2.5}$, the emission of $PM_{2.5}$ should not be higher than 10 $\mu p/m^3$ as an annual mean (who.int). This restriction will be legally binding in 2015 (Ibid.). The Asphekom project came to the conclusion that $PM_{2.5}$ was the most lethal of the pollutants investigated and that chronic exposure to the pollutant increases the risk of cardiovascular and respiratory diseases (Pascal et. al.). The largest burden on public health caused by one of the pollutants in the study was the burden caused by $PM_{2.5}$ (Ibid.). They found that if the cities in question would comply with the WHO-AQG guidelines, the life expectancy at 30 years of age and up would increase by up to 22 months (Ibid.). The only city in the study that at the time of the investigation did comply with the guidelines was Stockholm (Ibid.). A study of the short-term effects of coarse particles in the daily mortality in Stockholm (Meister et. al. 2012) due however show that $PM_{2.5-10}$ has significant effects on the mortality of the city's inhabitants and that this effect is higher during the winter months due to studded tires used because of the weather conditions (Ibid.).

In the same vein as the Correia et. al. (2013) article over the life expectancy changes in hundreds of US cities caused by changes in the $PM_{2.5}$ levels, Stroh et. al. (2005) investigated if socioeconomic status had any effect on the level of NO_2 in different municipalities in the Scanian county of Sweden. This study concluded that socioeconomic status variables should be included in analysis of the effects of air pollution since lower socioeconomic status covariates with higher levels of air pollutants (in this case higher levels of NO_2 (Ibid.)). They thus found that there is discrimination in the way that populations are exposed to air pollution and its adverse effects.

To summarize, the theory based on the characteristics of Sweden and on the effect of $PM_{2.5}$ emissions on life expectancy in the United States and in the European Union, is that fine particle pollution has a small but negative effect on life expectancy. In comparison to the US, Sweden has higher life expectancy and when looking at emissions of $PM_{2.5}$ and not concentrations in the air, Sweden has

lower levels of emissions. Since life expectancy is already higher and $PM_{2.5}$ is lower, the effect is expected to be smaller since the higher the life expectancy the more difficult it should be to affect it.

For cause-specific mortality, the effect of $PM_{2.5}$ is theorized to be in the opposite direction of the effect that the pollution has on life expectancy since mortality is the opposite of life expectancy. The theory is, therefore, that $PM_{2.5}$ has a small but positive effect on cause-specific mortality, ceteris parabus.

Method:

The hypothesis is that life expectancy is negatively correlated with fine particle pollution ($PM_{2.5}$); increases in particle pollution lead to decreases in life expectancy, hypothetically. Vice versa is expected to be true of the effect of PM_{25} on cause-specific mortality. Fine particle pollution ($PM_{2.5}$) should thereby also be positively correlated with mortality The factors that affect life expectancy and fine particle pollution have to be represented, as good as possible, by variables. The data that would best represent the different factors may not be available so the next best alternative needs to be found. The content of fine particle pollution in the air would be best represented by a measure of $\mu g/m^3$ and with yearly or daily averages. This data is however not available for enough counties or for a long enough time period to get an as large as possible number of observations and increase the probability of a statistically significant result of a regression model. An alternative measure has thus to be found.

Time Period:

The time period chosen in this analysis depends on the years of data available for the fine particle ($PM_{2.5}$) levels in the 21 Swedish counties. Data from the Swedish Environmental Protection Agency, SEPA, (Naturvårdsverket) is available for all counties in the years 1990, 2000, 2005-2011.

Area:

With the exception of gross regional product (used as a measure for income in the analysis), all data is available for individual municipalities. The areas of choice are however the 21 Swedish counties. The choice of the larger geographical areas is due to the lack of gross regional product for municipalities and because of the lack of time for calculating life expectancy for all 290 Swedish municipalities for all years used in the analysis.

In 1990 Sweden was divided into 25 different counties, but in 1997 Skåne county was formed by a merger between the Malmöhus and Kristianstad counties and in 1998 Västragötaland county was formed by a merger between Göteborg and Bohus, Skaraborg and Älvsborg counties (with the exception of two municipalities that joined Jönköping county)(scb.se, 2014). The data for 1990 is divided into the 21 counties that were formed in 1997/1998 for consistency. The data is available in this form from the Swedish Bureau of Statistics, SCB and when not it is corrected manually by subtracting these municipalities from one county and adding them to another.

Data:

The analysis is inspired by an article by Correia et. al. (2013) that uses US data to investigate the relationship between $PM_{2.5}$ and life expectancy in several hundred US counties. The variables used in that study are used in this analysis with some modification. Since Sweden is a more homogenous country ethnically no demographic variable for ethnicity is used and the variables for lung cancer and COPD deaths are combined into one variable. Both life expectancy at birth and life expectancy at age 40 are used as dependent variables, though separately. In addition, cause-specific mortality is used as a dependent variable in a third regression model to investigate the effect of pollution on the mediator variable, cause of death.

Life Expectancy:

Statistics Sweden offers numbers for life expectancy only on the national level. It does have numbers of deaths and of populations readily available for all regions yearly. The data is available for age-specific deaths and age-specific populations per year. Using this data, life expectancy per year and per county is calculated using Excel and the life table approach (see appendix for an example and a more detailed explanation of the method used). As mentioned under the heading "Area" above, it is possible to calculate life expectancy for all 290 municipalities using the life table method. To do so would, however, be much more time consuming and together with the lack of gross regional product on a municipality level, the choice was made to use the 21 Swedish counties instead.

This significantly reduces the number of observations and may come with additional problems for the statistical model and analysis.

$PM_{2.5}$:
The data for fine (2.5 μm) particle matter is available for all counties for the years 1990, 2000, 2005-2011 from the SEPA. The measure used is however emissions in tons per year and not the average daily $\mu g/m^3$ which has been the standard in previous research. Numbers in that measure are available for some municipalities for some years from the same agency but since the availability is inconsistent the choice was made to use emissions in tons per year instead.

Smoking:
Data for smoking for the population is not available. Statistics Sweden has data for how much income is spent on tobacco products per year on average in different counties but this information is not available for all years needed. As mentioned in the theory, the data for the year 2000 show that consumption of tobacco products is higher in more rural areas than in urban areas and that it is higher in the North of Sweden than in the South. There are also data available on the difference in income spent on tobacco products between blue-collar and while-collar workers with blue-collar workers spending more income than white-collar workers. This would strengthen the argument that higher educated people smoke less than lower educated people. The National Board of Health and Welfare (Socialstyrelsen) has data for the percentage of pregnant women that smokes three months before getting pregnant, early in the pregnancy and late in the pregnancy, but this only represents a small portion of the population and an estimate would need to be calculated to use this as a sample for the population as a whole. Previous studies of smoking and affect on mortality have used level of education as an indicator of socioeconomic class and deduced that people belonging to a lower socioeconomic class are more likely to smoke than people belonging to a higher socioeconomic class. More educated people are more likely to belong to a higher socioeconomic class and thus less likely to be smokers, ceteris parabus. The data used in this analysis is the percentage of the population

that hold at least a high school diploma and the percentage is calculated by using the total population between 15-74 in a county in the year in question and the population between 16-74 with at minimum a high school diploma. The difference in the population groups used is due to the difference in reporting of these statistics at Statistics Sweden and is kept in these groups for simplicity.

Cause of Death:
The cause of death used is cardiopulmonary or respiratory diseases that may be caused by air pollution. The variable is per 100,000 of population and the data is found at Statistics Sweden (for the year 1990) and The National Board of Health and Welfare (all other years). This variable is used both has a mediator independent variable in the regression models using life expectancy and as a dependent variable in a third regression model.

Income:
The kind of income used in the analysis is gross regional product. The data is available at Statistics Sweden for all years but 1990, however the first year available, 1993 is used instead. The numbers are adjusted for inflation using a consumer price index available at Statistics Sweden. 1990's CPI is used with gross regional product from 1993. Between 1990 and 1993 there was a financial crisis in Sweden so the income level in 1990 was probably higher in most counties than it was in 1993. This may affect the regression output but it is still the next best alternative found.

Population Change:
The change in population is calculated by using the total population of a county in year t and the total population in the year before, t-1. The data on population is available from Statistics Sweden.

Descriptive Statistics:

The means, the highest and the lowest values for the dependent variables used in the regression analysis (life expectancy at birth, life expectancy at age 40 and cause of death), for the independent variables used in the regression analysis (education, $PM_{2.5}$, income, population change) as well as for one of the weights used in the analysis (population density) are shown in the tables below. In addition, the same descriptive statistics are shown for population (in absolute terms).

For the highest/lowest values the relevant county is indicated in parenthesis.

The last two rows in each table show the means for the counties with high respective low population density. The two groups are formed using the average population density in the year in question, the values for which are shown below:

Table 1: Population Density in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	41.1	42.6	43.6	44.1	44.5	44.9	45.4	45.4	46.4
Highest	254.9 (Stockholm)	279.7 (Stockholm)	289.9 (Stockholm)	294.2 (Stockholm)	299 (Stockholm)	303.9 (Stockholm)	309.7 (Stockholm)	315.1 (Stockholm)	320.5 (Stockholm)
Lowest	2.7 (N-Botten)	2.6 (N-Botten & Jämtland)	2.6 (N-Botten)	2.6 (N-Botten)	2.6 (N-Botten)	2.5 (N-Botten)	2.5 (N-Botten)	2.5 (N-Botten)	2.6 (N-Botten)

Table 1: Population Density in Sweden, source: scb.se

Table 2: Life Expectancy at Birth in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	77.8	79.8	80.7	80.7	81.2	81.3	81.2	81.7	81.8
Highest	79.0 (Halland)	81.1 (Uppsala)	81.6 (Uppsala)	82.5 (Uppsala)	82.3 (Kronoberg)	82.6 (Halland)	85.0 (Jämtland)	82.8 (Halland)	83.0 (Kronoberg)
Lowest	76.8 (V-Botten)	78.3 (Gotland)	79.8 (Gävleborg)	77.6 (Gävleborg)	80.0 (N-Botten)	80.2 (N-Botten)	78.4 (Gävleborg)	80.6 (V-Norrland)	80.7 (N-Botten)
Density, high	78.1	80.3	81.1	81.1	81.4	81.6	81.5	82.1	82.2
Density, Low	77.7	79.7	80.6	80.5	81.1	81.2	81.0	81.6	81.6

Table 2: Life Expectancy in Sweden, source: scb.se (life expectancies calculated using a life table in Ex

The means, the highest and the lowest values for life expectancy at birth increase from 1990 to 2011, as seen in the table above. Although there is an overall increase (the life expectancy in 2010 is higher than in 1990), there is not a consistent increase nor are there no dips. The counties that are more populated consistently have higher life expectancy than the less populated ones. Note that the mean values for life expectancy at birth do not correspond to those same values for the country as a whole (e.g. in tables calculated by Statistics Sweden). One of the reasons for this is that the average age at death for an infant is calculated by using 0.1, for simplification, in these calculations and the number used by Statistics Sweden is unknown.

Table 3: Life Expectancy at Age 40 in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	39.3	40.9	41.7	412.0	42.1	42.3	42.1	42.6	42.8
Highest	40.5 (Halland)	42.1 (Uppsala)	42.7 (Skåne)	43.3 (Uppsala)	43.2 (Halland)	43.4 (Halland)	45.4 (Jämtland)	43.6 (Halland)	43.8 (Kronoberg)
Lowest	38.3 (V-botten)	39.9 (Gotland)	40.8 (V-Botten)	41.1 (Gävleborg)	41.3 (V-Botten)	41.3 (N-Botten)	39.3 (Gävleborg)	41.5 (V-Norrland)	41.7 (V-Norrland)
Popdensity, high	39.6	41.3	42.1	42.2	42.3	42.4	42.5	42.9	43.0
Popdensity, low	39.2	40.8	41.6	41.9	42.1	42.2	41.9	42.5	42.7

Table 3: Life Expectancy at age 40 in Sweden, source: scb.se (life expectancies calculated by life table in Excel)

Life expectancy at age 40 has increased from the first year, 1990, to the last year, 2011, but just as in the table above with life expectancy at birth, there are dips in the results and no steady increase. The counties with the highest or the lowest life expectancy at age 40 do not always correspond to the counties with the highest or lowest life expectancy at birth. This does not need to be the case since any age group can, and does, influence the calculation of life expectancy (a sample of the life table used for the calculations can be found in the appendix). Life expectancy at age 40 is higher in high population density counties than in low population density counties.

Table 4: Education (% with at least High School Diploma) in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	0.263	0.393	0.450	0.458	0.467	0.477	0.488	0.500	0.513
Highest	0.397 (Stockholm)	0.522 (Stockholm)	0.563 (Stockholm)	0.568 (Stockholm)	0.576 (Stockholm)	0.584 (Stockholm)	0.592 (Stockholm)	0.601 (Stockholm)	0.612 (Stockholm)
Lowest	0.216 (Kalmar)	0.339 (Kalmar)	0.401 (Kalmar)	0.410 (Kalmar)	0.421 (Kalmar)	0.431 (Gävleborg)	0.441 (Gävleborg)	0.454 (Gävleborg)	0.466 (Gävleborg)
Popdensity, high	0.286	0.429	0.484	0.485	0.494	0.504	0.514	0.533	0.538
Popdensity, low	0.251	0.382	0.439	0.447	0.456	0.467	0.478	0.490	0.503

Table 4: Education in Sweden, source: scb.se (percentages calculated in Excel)

The table above shows that the level of education of the population (a more specified explanation of the education variable is found in the methods section of this essay) rises during the observed years. The highest educated county is consistently Stockholm, whereas the lowest educated county changes from Kalmar to Gävleborg in 2008 (the values for 1991-1999 are not known since those years are not used in this analysis). High population density counties are more educated than low population density counties, which is not a surprise since metropolitan areas such as Stockholm (consistently the highest educated county) belong to this group.

Table 5: $PM_{2.5}$ Emissions in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	2194	1800	1813	1785	1782	1748	1710	1771	1780
Highest	6265 (Västergötland)	4249 (Stockholm)	4578 (Stockholm)	4614 (Stockholm)	4688 (Stockholm)	4619 (Stockholm)	4719 (Stockholm)	4883 (Stockholm)	4813 (Stockholm)
Lowest	621 (Jämtland)	507 (Jämtland)	559 (Jämtland)	569 (Jämtland)	563 (Jämtland)	5723 (Jämtland)	575 (Jämtland)	574 (Jämtland)	568 (Jämtland)
Popdensity, high	2778	2887	2932	2567	2590	2601	2619	3032	2627
Popdensity, low	1902	1460	1464	1472	1458	1406	1347	1377	1441

Table 5: PM_{25} Emissions, source: naturvardsverket.se

The mean level of $PM_{2.5}$ pollution per year has decreased in the years observed, however the highest values increase during the 2000s and the lowest values are more volatile. The values for high/low population density are lower in 2011 than in 1990, but there are slight dips and pikes in the years between. The levels are, however, consistently lower in low density areas.

Table 6: Income (Gross Regional Product)

Year	1990**	2000	2005	2006	2007	2008	2009	2010	2011
Mean	79.5	85.1	98.0	102.9	105.2	104.6	98.8	158.9	106.4
Highest	122.4 (Stockholm)	132.4 (Stockholm)	153.0 (Stockholm)	156.6 (Stockholm)	162.1 (Stockholm)	159.1 (Stockholm)	160.9 (Stockholm)	204.03 (Stockholm)	164.4 (Stockholm)
Lowest	69.4 (Sörmland)	72.5 (Gotland)	85.9 (Gotland)	88.7 (Gotland)	90.5 (Gotland)	87.2 (Gotland)	83.6 (Sörmland)	133.6 (Gotland)	91.5 (Blekinge)
Pop density, high	84.7	95.1	109.5	111.2	114.6	111.8	107.7	171.1	112.5
Pop density, low	81.9	81.9	94.4	99.5	101.4	101.7	95.2	155.1	103.9

Table 6: Gross Regional Product in Sweden, source: scb.se

** uses 1993 gross regional product with 1990's CPI

Table 7: Population Change (Percentage) in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	0.007	-0.001	0.015	0.003	0.004	0.004	0.005	0.004	0.003
Highest	0.155	0.110	0.009	0.051	0.164	0.016	0.019	0.017	0.018
Lowest	0.003	-0.01	-0.003	-0.49	-0.005	-0.003	-0.003	-0.002	-0.003
Pop density, high	0.009	0.005	0.006	0.000	0.008	0.008	0.01	0.009	0.007
Pop density, low	0.006	-0.003	0.000	0.005	0.002	0.003	0.003	0.003	0.001

Table 7: Population Change, source: scb.se

Table 8: Cause-Specific Mortality ("Cause of Death") in Sweden 1990-2011

Year	1990	2000	2005	2006	2007	2008	2009	2010	2011
Mean	703.8	608.6	489.2	504.6	503.4	507.8	499.3	522.9	515.2
Highest	824.8 (Värmland)	720.8 (Halland)	601.6 (V-Norrland)	660.4 (V-Norrland)	624.0 (Jämtland)	609.9 (V-Botten)	649.2 (N-Botten)	634.6 (Örebro)	637.9 (Kalmar)
Lowest	543.1 (Uppsala)	418.3 (Uppsala)	295.0 (Uppsala)	360.3 (Sörmland)	345.8 (Uppsala)	357.4 (Stockholm)	328.1 (Stockholm)	330.0 (Stockholm)	316.1 (Stockholm)
Popdensity, high	659.0	641.7	459.5	501.9	486.9	491.4	461.2	498.4	462.2
Popdensity, low	726.2	598.2	498.5	505.7	510.0	514.3	514.6	428.5	536.4

Table 8: Cause-Specific Mortality, source: scb.se (1990) and socialstyrelsen.se (2000, 2005-2011)

Outline of Statistical Models:

To see if changes in the independent variables have any effect on the dependent variable two types of regressions and several variations of the two are used. Since there are two gaps in the available data (1990 to 2000, 2000 to 2005), the use of a fixed effects regression model using panel data is limited to the latter years (2005 to 2011). A fixed effects model could be used with the years 1990, 2000 and 2010, to keep the gaps between the years consistent, but the number of observations would then be smaller which may affect the results for the worse. A cross-sectional ordinary least square, OLS, regression model is used for each of those years instead to compare the three.

There are three different dependent variables used in the analysis: 1) life expectancy at age 40; 2) life expectancy at birth; and 3) deaths due to lung and/or cardiopulmonary diseases per 100,000 of the population. In vein of the Correia et. al. (2013) article, the cause of death variable is also used as an explanatory variable in the regressions with life expectancy as the dependent variable. Life expectancy at age 40 is the main model since air pollution damages health through long-term exposure.

For all statistical analyses, the software program Stata is used. For calculations of life expectancy, education and others, however, Excel is used.

Linear Multivariate Regression:

For the years 1990, 2000 and 2010, the data used is cross-sectional and not panel data and regressions are run separately for each year. There are three different versions for each year of the full OLS model in this analysis, one where the dependent variable is life expectancy at age 40 (the main version), one where the dependent variable is life expectancy at birth and one where the dependent variable is deaths per 100,000 where the cause of death is related to respiratory and/or cardiopulmonary disease.

The first version is life expectancy at age 40, where the equation for the full model is as follows:

$$le40_i = \propto + \beta_1 pm2.5_i + \beta_2 educ_i + \beta_3 inc_i + \beta_4 popchange_i + \beta_5 cod_i + u_i$$

where

le40 is life expectancy at age 40, in years

pm2.5 is the level of fine particle pollution in tons per year
educ is the percentage of the population between 16-74 with at least a high school education
inc is the gross regional product
popchange is the change in population from the previous year, in percent
cod (cause of death) is deaths per 100,000 of the population due to causes of death related to lung- and/or cardiopulmonary diseases.
u is the residual

The second version is life expectancy at birth, the equation is

$$le_i = \propto + \beta_1 pm2.5_i + \beta_2 educ_i + \beta_3 inc_i + \beta_4 popchange + \beta_5 cod_i + u_i$$

where
le is the life expectancy at birth, in years
the remaining variables are as defined above

The third version is cause of death, the equation as follows:

$$COD_i = \propto + \beta_1 pm2.5_i + \beta_2 educ_i + \beta_3 inc_i + \beta_4 popchange_i + \beta_5 cod_i + u_i$$

where
COD is the cause of death (for lung/cardio/pulmonary disease) per 100,000 of population
the remaining variables are as defined above.

For each version of the OLS model, an analysis of the effects of the independent variables on the dependent variable is done. Linear regressions are run for each individual year by using Stata and the command *reg* with the dependent variable followed by the explanatory variables. The first linear regression is run with only PM_{25} as an explanatory variable to see if the amount of pollution solely has any effect on the dependent variable. The main version of the regression model uses all independent variables defined above (PM_{25}, education, income, population change and cause of death) except for when cause of death is the dependent variable. All regressions use *robust standard errors* to account for any heteroskedasticity in the error terms. The results of interest are R^2 (to see how much of the variation in the explained that is accounted for by the explanatory variables), the coefficients of the explanatory variables and the constant (to see how much and in what direction the variables affect the dependent variable) and the p-values (to see if the

variables and the constant are statistically significant, in this case at the 5% level). Since all regressions use *robust standard errors*, adjusted R^2 is not calculated and it is therefore not possible to discern if the added regressors are "good" or "bad" variables.

To see if the population density or size of the counties contributes to the effect of the independent variables on the dependent variable, weights are used. The first weight is the square root of the population density and the second is the inverse of the county area (the information for which is from 1999). The Correia et. al. (2013) article used similar weights in their calculations.

The results from the years 1990, 2000 and 2010 are compared to see if there is any difference in the effect the explanatory variables have on the three dependent variables: any difference in the direction and/or magnitude of the coefficients, any change in R^2 and any difference in what variables show statistically significant result are considered.

Fixed Effects Model:
There are three different versions of the full fixed effects model in this analysis, one where the dependent variable is life expectancy at birth, one where the dependent variable is life expectancy at age 40 and one where the dependent variable is deaths per 100,000 where the cause of death is related to lungand/ or cardiopulmonary disease. The first version is life expectancy at birth, where the equation for the full model is as follows:

$$le40_{it} = \propto + \beta_1 pm2.5_{it} + \beta_2 educ_{it} + \beta_3 inc_{it} + \beta_4 popchange_{it} + \beta_5 cod_{it} + \delta_1 2006 + \delta_2 2007 + \delta_3 2008 + \delta_4 2009 + \delta_5 2010 + \delta_6 2011 + v_{it}$$

where
le40 is life expectancy at age 40, in years
pm2.5 is the level of fine particle pollution in tons per year
educ is the percentage of the population between 16-74 with at least high school education
inc is the gross regional product
popchange is the change in population from the previous year, in percent
cod (cause of death) is deaths per 100,000 of the population due to causes of death related to lung- and/or cardiopulmonary diseases.

2006-2011 are dummy variables for difference intercept/constant for the different years. 2005 is the base year and thus therefore not have its own dummy

v is the residual (the composite error $v_{it} = a_i + u_{it}$ where a_i is the fixed effect)

The second version is life expectancy at age 40, the equation is

$$le0_{it} = \propto + \beta_1 pm2.5_{it} + \beta_2 educ_{it} + \beta_3 inc_{it} + \beta_4 popdensity_{it} + \beta_5 cod_{it} + \delta_1 2006 + \delta_2 2007 + \delta_3 2008 + \delta_4 2009 + \delta_5 2010 + \delta_6 2011 + v_{it}$$

where

le0 is the life expectancy at birth, in years

the remaining variables are as defined above

The third version is cause of death, the equation as follows:

$$COD_{it} = \propto + \beta_1 pm2.5_{it} + \beta_2 educ_{it} + \beta_3 inc_{it} + \beta_4 popchange_{it} + \delta_1 2006 + \delta_2 2007 + \delta_3 2008 + \delta_4 2009 + \delta_5 2010 + \delta_6 2011 + v_{it}$$

where

COD is the cause of death (for lung/cardiopulmonary disease) per 100,000 of population

the remaining variables are as defined above

To run the regressions for the fixed effects model in Stata the command *xtreg* is used with *fe* to indicate fixed effects. With the exception of this, the different versions of the regression models and the interpretations of the results is similar to the cross-sectional model and is outlined in that section of the essay above.

Tests:

For both the OLS model and the fixed effects model the regressions are run accounting for heteroskedasticity in the error terms by using the robust standard errors option in Stata. The regressions are also run with two different kinds of weights: the square root of the population density to see if the difference in the density of the different counties affects the results and the inverse of the area of the counties to see if the geographical size of the county affects the results. Those two weights were chosen due to their use in the Correia et. al. (2013) article on which this analysis is inspired. The data for the areas is the same throughout all the years (although as aforementioned, in 1990 there were 25 counties)

and is from 1999. For the population density, since the weight needs to be constant for each county in the regression, the value for the year 2005 is used for the fixed effects model for all the years, whereas the value for the years 1990 (the values for 1991 are used as proxies for 1990 due to lack of data), 2000 and 2010 is used in each of the OLS regressions respectively.

Variants:

In the chapter that follows the results from the different regressions are shown in up to seven different variants called models. For the regressions where any of the two life expectancy variables are the dependent the seven models are as follows:

1. $PM_{2.5}$ on Life Expectancy (at 40/birth)
2. $PM_{2.5}$ and Education on Life Expectancy
3. $PM_{2.5}$, Education and Income on Life Expectancy
4. $PM_{2.5}$, Education, Income and Population Change on Life Expectancy
5. $PM_{2.5}$, Education, Income, Population Change and Cause of Death on Life Expectancy (the main version)
6. Models 5 (see above) but with the Square Root of the Population Density as an analytical weight
7. Model 5 (see above) but with the Inverse of the Land Area as an analytical weight

When, however, cause of death is the dependent variable it cannot also be used as an explanatory variable so the six models are as follows:

1. $PM_{2.5}$ on Life Expectancy (at 40/birth)
2. $PM_{2.5}$ and Education on Life Expectancy
3. $PM_{2.5}$, Education and Income on Life Expectancy
4. $PM_{2.5}$, Education, Income and Population Change on Life Expectancy (the main version)
5. Models 4 (see above) but with the Square Root of the Population Density as an analytical weight
6. Model 4 (see above) but with the Inverse of the Land Area as an analytical weight

Results from Statistical Models:

The results for the main model with life expectancy at age 40 as the dependent variable and the two complementary models (life expectancy at birth and cause of death as the dependent variable, respectively) are presented below. All versions of the regression model use fine particle pollution as the main independent variable and with the exception for the first version of all models use a number of control variables. The main version uses fine particle pollution and the following controls: level of education, level of regional income, change in population from the previous year and cause of death (cardiopulmonary or lung diseases). The level of significance is 5% in all versions and the number of decimals is set to four for all models to be able to capture the small effect of fine particle pollution on life expectancy and cause of death. In the regression output an asterisk (*) denotes that the coefficient is statistically significant. The robust standard errors are in parenthesis below the coefficient for each variable.

Table 9: Summary Statistics for 1990

The results for the first year, 1990, are presented below. From the descriptive statistics section above, the following means are found:

Life expectancy at birth (e0), in years	77.82
Life expectancy at age 40 (e40), in years	39.34
Fine particle pollution (pm2.5), in thousands of tons	2194.31
Education, in percentage of population with at least high school diploma	0.263
Gross regional income, adjusted for inflation	164.95
Change in population from previous year, in percent	0.007
Cause of death (COD), for cardiopulmonary and lung diseases per 100,000 of population	703.78

Table 9: Summary Statistics 1990, sources: scb.se (see details in "Descriptive Statistics" sub-section)

Table 10: OLS Regression 1990 with Life Expectancy at Age 40 as the Dependent Variable

e40	Model 1	Model 2	Model 3	Model 4	Model 5	Model 8	Model 9
$PM_{2.5}$	-0.0000	-0.0001	-0.0001	0.0000	-0.0000	0.0000	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Education		3.3034	5.4209*	-4.8586	-7.3564*	-7.0652*	-5.9098*
		(2.6850)	(2.1819)	(3.4569)	(3.2677)	(2.0733)	(1.8736)
Income			-0.0144	0.0088	0.0075	0.0024	0.0080
			(0.0092)	(0.0103)	(0.0095)	(0.0062)	(0.0068)
Population				137.8543*	100.4353*	88.4071*	81.2705
Change				(29.7143)	(38.3117)	(33.1928)	(40.9546)
COD					-0.0033	-0.0035	-0.0029
					(0.0017)	(0.0017)	(0.0023)
Constant	39.4362*	38.6480*	39.2118*	38.9096*	42.3256*	42.8884*	42.0313*
	(0.2120)	(0.6682)	(0.5651)	(0.3637)	(1.8778)	(1.8490)	(2.4892)
R^2	0.0157	0.0749	0.1167	0.4724	0.5471	0.6324	0.6332

Table 10: OLS with LE at 40, 1990

The first cross-sectional regression model shows that fine particle pollution is never statistically significant in 1990 nor is the direction of the coefficient consistently in one direction. There are plenty of explanations for why this is. One reason can be related to the configuration of the statistical model, the choice of data (or the availability of data) or the choice of statistical model. If one disregards the possibility of a problem with the statistical model, the result of the regression model may differ from what was expected due to how the pollution was distributed. In 1990, the South of Sweden had higher levels of emissions than the mid or the North of Sweden. A regression of the main model (model 5) was run with a categorical variable that indicated to which of the three parts of Sweden a county belong and while not included in the results, the output of that regression showed a statistically significant and negative effect of fine particle pollution on life expectancy at age 40. That result is in line with the theory: a higher level of fine particle pollution should, theoretically, lead to a lower level of life expectancy. Here, however, the main independent variable is not statistically significant and in the main model, model 5, the coefficient is zero. An explanation for this may be that the areas with the highest levels of emissions are also areas where the population density is higher, industrial activity is higher, education is higher and life expectancy is higher. Smoking, according to Swedish data from 2000, is more common in less populated areas and is higher in the North than in the South of Sweden. This factor may be one of the reasons for why life expectancy is lower in areas with lower levels of air pollution. This factor alone cannot, obviously, account for why life expectancy is lower in those areas but may be a probably cause.

An abnormality in the control variables is the changing direction of the education variable. Education is expected to have a positive effect on life expectancy; education as used as a measure of smoking prevalence should be positively correlated with life expectancy as non-smokers are expected to live longer than smokers, on average. Education in itself should also, in theory, be positively correlated with life expectancy as more educated people are expected to live longer. An explanation for why education in the latter versions of the regression model is negatively correlated with life expectancy can be that the areas that are more heavily polluted are also the areas where the average level of education is higher. The descriptive statistics sub-section above show that the high population density counties have higher levels of pollution and higher

levels of education. The higher education may thus represent the areas with more pollution and thereby be negatively correlated with life expectancy. The variable may capture the effect of pollution in this way. It is also possible that the result is skewed because of the small number of high population density counties and large number of low population density counties.

The rest of the control variables are behaving in the way that was expected: income could be either positively or negatively correlated with life expectancy since higher income may be indicative of a more productive industry and a better economy may be both good and bad for life expectancy. Higher income may lead to higher levels of pollution that may lead to lower life expectancy. Higher income may also lead to higher life expectancy. Population change was expected to be, and is, positively correlated with life expectancy. It is also statistically significant in all but one model (model 7). Its coefficient was expected to be positive since a positive change in population means that either mortality was lower than the birthrate, net-immigration was higher than the previous year or a combination. The dependent variable, life expectancy at age 40, affects population change in itself so population change can be seen as a confounding variable. The last control variable, cause of death, is a mediator between fine particle pollution and life expectancy since the level of pollution affects cause-specific mortality (respiratory and/or cardiopulmonary diseases) which cause of death represents. Cause-specific mortality then affects life expectancy and an increase in deaths due to air pollution (in this case) should lead to a decrease in life expectancy.

The standard errors used in this regression model, and all following, are robust standard errors. The use of robust standard errors means that adjusted R^2 is not included in the regression output. R^2, in contrast to adjusted R^2 , increases with the addition of new variables regardless if the variables are “good” or “bad” variables. The variables used in this, and the following, regression are included because of their relevance in previous research and the argument for their use is outlined in the theory and method sections above.

Table 11: OLS Regression 1990 with Life Expectancy at Birth as the Dependent Variable

e0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
$PM_{2.5}$	-0.0001	-0.0001	-0.0001	-0.0000	-0.0000	2.19x10^-6	-0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education		3.6799	6.1971*	-4.8243	-6.8613	-7.2900*	-5.2573
		(3.2947)	(2.4835)	(3.6617)	(3.9654)	(2.6864)	(2.9262)
Income			-0.01718	0.0077	0.0067	0.0002	0.0102
			(0.0115)	(0.0120)	(0.0117)	(0.0078)	(0.0106)
Population				147.8026*	117.2864*	98.7890*	91.5348
Change				(32.1920)	(49.4709)	(42.2647)	(53.3355)
COD					-0.0027	-0.0033	-0.0025
					(0.0026)	(0.0023)	(0.0031)
Constant	78.0122*	77.1341*	77.8043*	77.4804*	80.2661*	81.4350*	79.9562*
	(0.2529)	(0.8149)	(0.7478)	(0.5437)	(2.8450)	(2.5613)	(3.4965)
R^2	0.0444	0.0986	0.1423	0.4443	0.4809	0.5637	0.5498

Table 11: OLS with LE at birth, 1990

Fine particle pollution, while never statistically significant, has a negative effect on life expectancy at birth in the majority of the variants of the regression model. All the coefficients of the control variables are in the same direction and of similar size as the independent variables in the regression model with life expectancy at age 40 as the dependent variable (see above). The statistical significance of the control variables, on the other hand, is not always the same as in the regression above. As laid out above, the explanation for some of the abnormalities in the results can be because the higher levels of pollution and life expectancy (as well as higher levels of some of the control variables) are concentrated in the South of Sweden. This might explain why education, a variable that because of its definition should have a positive effect on life expectancy, is negative in the latter versions of the models, including the main version, model 5. It may be that this variable carries the effect of pollution with it since the high pollution areas are also the high education areas. Unexplained factors of higher life expectancy can be the reason why the areas with high levels of emissions are still areas with lower mortality. When accounting for the population density or land area, however, education still has a negative effect on life expectancy. A factor that is explained above is smoking prevalence: tobacco consumption is higher in the North of Sweden and it is higher in rural areas than in urban areas. Therefore, metropolitan areas such as Stockholm, Gothenburg and Malmö where there are more fine particle pollution, smoking is less common and that might mean that people have fewer preexisting conditions that would be aggravated by the air pollution and therefore life expectancy can still be high.

Table 12: OLS Regression 1990 with Cause of Death as the Dependent Variable

COD	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$PM_{2.5}$	-0.0147	0.0005	-0.009	-0.0082	-0.0004	0.0052
	(0.0076)	(0.0084)	(0.0087)	(0.0095)	(0.0075)	(0.0069)
Education		-1.37x10^3*	-1592.584*	-752.2535	-571.6231	-561.2489*
		(266.8883)	(200.0988)	(376.3739)	(300.8663)	(198.1031)
Income			1.5136	-0.3855	-1.1884	-1.4155*
			(07919)	(1.113)	(0.8729)	(0.5773)
Population				-11269.31*	-12815.26*	-13492.8*
Change				(4216.684)	(2901.986)	(1679.232)
Constant	736.0182*	1063.1138*	1004.0614*	1028.763*	1036.312*	1045.093*
	(26.7868)	(69.2213)	(55.7295)	(41.7699)	(31.068)	(21.1389)
R2	0.0787	0.5529	0.5742	0.6849	0.7905	0.8593

Table 12: OLS with Cause of Death, 1990

In theory, the direction of the coefficient of PM_{25} in the model where cause of death (cause-specific mortality due to respiratory and/or cardiopulmonary diseases) is the dependent variable should be opposite of what is expected in the models using life expectancy as the explained. This means that the direction of the effect of fine particle pollution should, in theory, be positive since larger amounts of emissions should lead to higher cause-specific mortality. In the regression output for 1990 above, the result is the opposite for all versions of the model except for 2 and 6. The coefficients are statistically insignificant in all versions. The explanation for the opposite effect of PM_{25} than expected from the previous two regression models using life expectancy might be reiterated here. A reason for the strange results could be that cause-specific mortality is lower in counties with higher life expectancy. Unexplained factors and control variables such as education and population change may be what cause the negative effect of PM_{25} on cause of death. The fact that the cause-specific mortality is lower in high population density (and high pollution) counties can also be a reason for why the effect is contrary to theory.

The control variables are behaving as expected according to the regression output. Education, in theory should be negatively correlated with cause of death. If representing smoking prevalence, people with higher education are less likely to smoke and thus higher education should have a negative impact on cause-specific mortality. In contrast o how education behaved in the regressions with life expectancy, in this regression output it consistently affects cause of death in the expected direction. This does not mean that the potential explanation for the curious results of education in the life expectancy regressions must be false. It may mean that education, in this regression model, does not absorb the effect of pollution in the same way as it did with life expectancy. The dependent variables are also each other's opposite; the dependent variables may all deal with deaths, but life expectancy puts more emphasis on who lived through a period than on who died and vice versa for cause-specific mortality.

Table 13: Summary Statistics for 2000

Summary statistics for the year 2000:	
Life expectancy at birth (e0), in years	79.85
Life expectancy at age 40 (e40), in years	40.95
Fine particle pollution (pm2.5), in thousands of tons	1800.14
Education, in percentage of population with at least high school diploma	0.39
Gross regional income, adjusted for inflation	228
Change in population from previous year, in percent	-0.001
Cause of death (COD) for cardiopulmonary and lung diseases per 100,000 of population	608.57

Table 13: Summary Statistics for 2000, source: various (for more details see "Descriptive Statistics" sub-section above)

Table 14: OLS Regression 2000 with Life Expectancy at Age 40 as the Dependent Variable

e40	Model 1	Model 2	Model 3	Model 4	Model 5	Model 8	Model 9
$PM_{2.5}$	-0.0001	-0.0002*	-0.0002	-0.0003*	-0.0003*	-0.0002	-0.0003*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education		8.6423*	8.3163*	4.5077	4.2045	3.7619	9.6849
		(2.5430)	(2.9173)	(3.1670)	(4.5886)	(5.7119)	(5.9755)
Income			0.0028	0.00206	0.0025	-0.0065	-0.0013
			(0.0106)	(0.0103)	(0.0111)	(0.0700)	(0.0112)
Population				47.9221*	46.4702	38.2095	36.0384
Change				(18.3297)	(24.3296)	(28.4453)	(41.093)
COD					-0.0002	-0.0010	-0.0010
					(0.0028)	(0.0014)	(0.0043)
Constant	41.0633*	37.9628*	37.8665*	39.5187*	39.7585*	40.7664*	37.2822*
	(0.1744)	(0.9205)	(1.0560)	(1.4053)	(3.0957)	(1.2811)	(4.2187)
R^2	0.0197	0.4149	0.4171	0.5235	0.5237	0.4727	0.6187

Table 14: OLS with LE at 40, 2000

In contrast to the regressions run with 1990's data, $PM_{2.5}$ is found to be statistically significant in some of the versions using 2000's data. When using life expectancy at age 40 as the dependent variable, $PM_{2.5}$ is statistically significant in models 2,4,5 and 7. The coefficients are consistently negative, which means that the effect is what was expected. Fine particle pollution should, in theory, have a negative effect on life expectancy, as it is harmful to health. So why should the results of the regression model in 2000 but not in 1990 show effects of fine particle pollution that is in the expected direction? Though the counties that include the large metropolitan areas (Stockholm, Gothenburg and Malmö) still have high levels of pollution and high levels of life expectancy, education and income, the distribution of PM_{25}/life expectancy is smaller in 2000 than it is in 1990. The PM_{25}/life expectancy does not, however, show a downward trend; a higher level of pollution does not necessarily mean a lower level of life expectancy. There are also counties where life expectancy is significantly lower than the Swedish average yet the pollution levels are lower than average (an example is Norrbotten). An explanation for this phenomenon could be that certain municipalities have high levels of pollution but the majority of the municipalities in the county do not. Regardless, the more even distribution of PM_{25}/life expectancy may contribute to the statistically significant and negative effect of the main explanatory variable on the explained.

In terms of the control variables, education behaves in the expected way. Education has a positive effect on life expectancy, which is what was hypothesized in the theory section above. Income is never statistically significant but the sign of the coefficient is inconsistent. As theorized, income could have either a positive or a negative effect on life expectancy. An increase in income could carry with it an increase in the level of pollution which may be damaging to health and thus have a negative effect on life expectancy. It is, however, only negative when used in the versions of the model weighted by either the population density or the land area of the counties. Thus, a higher population density or a smaller geographical area may contribute to the negative effect.

In contrast to the regression with life expectancy at age 40 using 1990's data, population change is only statistically significant in one version of the model. The magnitude of the coefficient is also lower than in 1990, though the direction is the same.

It could be that other variables carry more weight than population change. The positive effect of education, for example, can be one of the disparities between 1990 and 2000 that changes the effect of population change. Lastly, cause of death is never statistically significant but the effect on life expectancy is negative which is what was theorized.

Table 15: OLS Regression 2000 with Life Expectancy at Birth as the Dependent Variable

e0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	-0.0001	-0.0001	-0.0001	-0.0002	-0.0002	-0.0001	-0.0005*
PM_{25}	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Education		9.5078* (3.1315)	5.9616	2.3237	0.9364	2.9575	14.6715
			(4.1290)	(5.1650)	(7.0163)	(8.5388)	(9.3826)
Income			0.0076	0.0069	0.0087	-0.0047	0.0004
			(0.0162)	(0.0164)	(0.0155)	(0.0086)	(0.0156)
Population				45.7754	39.1331	30.9337	48.8448
Change				(26.4153)	(29.7684)	(36.4651)	(57.2165)
COD					-0.0011	-0.0011	0.0039
					(0.0035)	(0.0049	(0.0064)
Constant	79.9503*	76.5322*	77.2114*	78.7895*	79.8866*	80.1953*	72.5275*
	(0.2139)	(1.1040)	(1.5896)	(2.2855)	(4.6589)	(5.8086)	(16.6064)
R^2	0.0103	0.3495	0.2168	0.2970	0.3006	0.2880	0.5495

Table 15: OLS with LE at birth, 2000

$PM_{2.5}$ is negatively correlated with life expectancy at birth in all models above but only statistically significant in model 7. The regression with life expectancy at birth does not follow the regression with life expectancy at age 40 in that regard. It is only statistically significant when the observations are weighted by the land area, which may account for the large quantities of air pollution in some of the smaller counties (for example Gotland).
When it comes to the control variables, it is curious that in the main version of the model (model 5), the coefficient of education is much lower than in the other variants. There may be a problem with the data or it can be explained by looking at the other independent variables. A large coefficient of income may dwarf the effect of education since in some ways the variables represents similar phenomenon; both income and education can represents a lower smoking prevalence since data from Statistics Sweden show that people living in larger cities consume less tobacco products than people living in rural areas (especially in the North). The variable of income can represent the difference between the counties with larger cities or metropolitan areas and counties that are mainly rural since the counties that are more urbanized tend to have larger gross regional products. The smoking prevalence can therefore be represented by the income variable as well as the education variable. It is still strange that education fluctuates so much in the different variants and it might be more probable that there is a data error or an error in the reporting in the table above versus the Stata output.

Income and population changer follow the same pattern as for life expectancy at age 40. Population change is still lower in the output for 2000 than in the output for 1990. Cause of death is negative, as expected, in two of the three versions where included but it is positive in the last version. In that version, the effect of air pollution of life expectancy is larger than in all other variants so that may account for why the mediator, cause-specific mortality, is behaving in the opposite way of what was presumed.

Table 16: OLS Regression 2000 with Cause of Death as the Explanatory Variable

COD	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$PM_{2.5}$	0.0294	0.0358	-0.0102	- 0.0081	0.0018	-0.0053
	(0.0154)	(0.0190)	(0.0087)	(0.0102)	(0.0072)	(0.0083)
Education		-341.6318	-1673.664	-1212.338*	-1182.204*	-1036.731*
		(315.3055)	(181.3337)	(258.1146)	(255.2971)	(253.5654)
Income			1.4839	1.5700*	1.2008*	1.6772*
			(0.7735)	(0.6220)	(0.5451)	(0.5600)
Population				-5804.752*	-6832.058*	-8232.467*
Change				(2328.034)	(2077.743)	(1779.722)
Constant	555.6936*	678.5119*	1158.841*	958.7182*	963.0949*	874.0501*
	(33.7588)	(110.4625)	(53.4374)	(80.6518)	(78.9886)	(80.4122)
R^2	0.1755	0.2030	0.7848	0.8540	0.9168	0.9082

Table 16: OLS with Cause of Death, 2000

In contrast to the regression model with cause-specific mortality as the dependent variable for 1990, the effect of PM_{25} is not consistently negative for 2000. In model 1, 2 and 5, the direction of the effect is positive which is the expected direction. The coefficient is never statistically significant. The explanation for the negative effect of fine particle pollution on cause-specific mortality above was that high pollution counties still have lower deaths per 100,000 caused by respiratory and/or cardiopulmonary disease than counties with lower levels of emissions. The reason could be that people tend to smoke less in counties with larger cities and metropolitan areas and non-smokers are, on average, less sensitive to air pollution though this may not always be the case.

Education, income and population change follow the theory in the regression output. Education has a negative effect on cause of death since as an indicator for smoking a higher level of education leads to a lower levels of deaths due to lung and/or cardiopulmonary illnesses. Income is an ambiguous variable and an effect in either direction can be explained. Higher income might indicate a higher level of industry in a county and depending on what kind of production, more production may lead to higher levels of emissions of air pollutants. Population change indicates a growing population (when positive), which means that either the population grew by births being higher than deaths or by net-immigration. The effect of population change should be in the opposite direction to what is expected in a regression of life expectancy. The effect in this case is negative, so it is in line with theory.

It may be worth mentioning that the R^2 values in this regression output is very high which is unexpected since the unobserved factors should account for a larger percentage of the variance in cause of death. It does not have to mean that there is something wrong with the model, but it is still strange that the values of the explained variance is around 80% or higher in the last three variants of the regression model.

2010:

For summary statistics for 2010, see the next section (fixed effects model).

Table 17: OLS Regression 2010 with Life Expectancy at Age 40 as the Dependent Variable

e40	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 9
$PM_{2.5}$	-0.0001 (0.0001)	-0.0002*	-0.0001	-0.0002*	-0.0002*	-0.0002*	-0.0002*
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education		9.0373*	9.8908*	1.4561	0.3941	2.7458	4.8036
		(2.5420)	(2.8157)	(2.7559)	(4.1398)	(4.6934)	(5.4881)
Income			-0.0047	-0.0024	-0.0019	-0.0044	-0.0050
			(0.0035)	(0.0035)	(0.0036)	(0.0028)	(0.0037)
Population				81.0015*	71.9459*	74.325	86.7663
Change				(27.3745)	(31.8458)	(37.2932)	(40.8555)
COD					-0.0012	-0.0001	0.0017
					(0.0022)	(0.0027)	(0.0028)
Constant	42.7359*	38.4684*	38.7246*	42.2668*	43.3563*	42.0501*	40.2916*
	(0.1860)	(1.1767)	(1.1442)	(1.3365)	(2.7602)	(3.2898)	(3.7040)
R^2	0.1093	0.3912	0.4201	0.6190	0.6238	0.6298	0.6651

Table 17: OLS with LE at 40, 2010

The results from the regression output of the statistical model with life expectancy at age 40 using 2010's data are the results that are most in line with the theory for the main explanatory variable, PM_{25}. The direction of the effect is consistently negative and it is statistically significant in all versions of the model but 1 and 3. That the results are more in line wit theory can be because the trend of PM_{25}/life expectancy is slightly downward and the distribution of the observations is smaller than for example 1990. The three counties with the large metropolitan areas still have high levels of air pollution and above average life expectancy, none of which is surprising. It is still problematic that the counties with the highest pollution levels are also the counties with high life expectancy, education and income, so one has to assume that unobserved factors and these variables have a large influence on the life expectancy even though the regression output does not indicate that education, income, population or cause of death have large effects on life expectancy nor are they statistically significant often. It is not surprising that this is the way it is but it may affect the regression model.

As in the regression model for 2000 with life expectancy at birth, the effect of education is low in the main version (5) of the model for 2010 with life expectancy at age 40 as the explained. Since this is now recurring, it seems less likely to be an input error. The decrease I the effect happens with the addition of first population change, in model 4, and then again with the inclusion of cause of death. These variables may therefore swallow some of the effect of education. Cause of death represents cause-specific mortality relating to respiratory and/or cardiopulmonary diseases. This factor is related to smoking, as is education.

Income has a consistently negative effect on life expectancy at age 40. As aforementioned, the sign of income can be either negative or positive without being curious since an increase in income can lead to higher life expectancy or lower life expectancy, the latter of which may happen because income increases lead to increases in emission levels through certain types of production.

Change in population affects life expectancy in the direction that is expected. The variable is something of a confounding variable since life expectancy also affects population change. Or at the very least, changes in mortality affect both population

change and life expectancy. Cause of death is, once again, positive in the last version of the model. Cause-specific mortality is not expected to have a positive effect on life expectancy since mortality and life expectancy are opposites. It may be that the air pollution carries the effect of the cause-specific mortality and this is why the effect is in the opposite direction.

Table 18: OLS Regression 2010 with Life Expectancy at Birth as the Dependent Variable

e0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 8	Model 9
$PM_{2.5}$	-0.0001 (0.0001)	-0.0002* (0.0001)	-0.0001	-0.0002*	-0.0002*	-0.0002*	-0.0002*
			(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education		9.9815* (2.9449)	11.2820*	1.3385	2.4260	4.4757	7.0233
			(3.2369)	(43.9820)	(5.2749)	(6.3227)	(7.6125)
Income			-0.0072 (0.0041)	-0.0023	-0.0049	-0.0047	-0.0063
				(0.0015)	(0.0033)	(0.0029)	(0.0044)
Population				89.6618*	95.8356*	87.1949	76.9648
Change				(29.1745)	(28.6453)	(41.4870)	(52.1208)
COD					-0.0007	0.0013	0.0024
					(0.0021)	(0.0030)	(0.0033)
Constant	81.8143*	77.1009*	77.4913*	81.4724*	80.8506*	79.6031*	78.1562*
	(0.2013)	(1.3573)	(1.3904)	(1.8728)	(3.3899)	(4.1410)	(4.8213)
R^2	0.0182	0.3905	0.4455	0.6498	0.6516	0.6129	0.5555

Table 18: OLS with LE at birth, 2010

PM_{25} has a consistently negative effect on life expectancy, which is in line with the theory. The magnitude of the effect is small, something that it has consistently been throughout the different years and which is expected since the form that the fine particle pollution data is in is thousands of tons. The variants of the model where fine particle pollution is not statistically significant are model 1 and 3, the same as in the regression for life expectancy at age 40.

Among the control variables, education dips sharply when population change is added to the model and then rises again with the addition of cause of death. It seems like population change affects life expectancy to a greater degree than education because when used together the effect of education falls and the coefficient of education becomes statistically insignificant. In some way, population change causes the effect that education has on life expectancy to change.

Income has a consistently negative effect on life expectancy and, as mentioned above, a negative coefficient of income can be in line with theory if one considers that an increase in income comes with an increase in heavy pollution production. Cause of death, on the other hand, is not expected to have a positive effect on life expectancy, yet when the variables are weighted by population density and land area respectively, cause-specific mortality has a positive effect on life expectancy. It may be so that when using these weights, the effects of high population density and small land area mean that higher numbers of deaths per 100,000 is correlated with higher life expectancy, something that is odd.

Table 19: OLS Regression 2010 with Cause of Death as the Dependent Variable

COD	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$PM_{2.5}$	-0.0248	-0.0002	-0.0051	-0.0026	0.0035	0.0013
	(0.0105)	(0.0067)	(0.0088)	(0.0067)	(0.0034)	(0.0039)
Education		-1557.905*	-1661.85	-880.9936*	-854.344*	-817.4577*
		(199.5726)	(182.2744)	(371.8678)	(286.1239)	(268.9057)
Income			0.5780	0.3614	0.3300	0.4358
			(0.4196)	(0.3989)	(0.1982)	(0.2327)
Population				-7498.915*	-8404.468*	-8904.574*
Change				(2864.342)	(2241.398)	(2065.18)
Constant	525.7184*	1261.3884*	1230.183*	902.2563*	877.0838*	860.0267*
	(24.8864)	(94.2877)	(93.3481)	(170.3221)	(128.5981)	(125.2108)
R^2	0.1891	0.7339	0.7550	0.8390	0.9218	0.9061

Table 19: OLS with Cause of Death, 2010

PM_{25} behaves in the opposite what of what is expected from theory; fine particle pollution is negatively correlated with cause-specific mortality in the regression output above but is expected to have a positive effect on mortality since higher levels of pollution should, logically, be correlated with higher numbers of death, ceteris parabus. The explanation of the high pollution, high life expectancy, high education and income counties, as used before, can also be applied here. When weighted by population density or land area, the effect of fine particle pollution is positive.

The control variables are expected to affect the dependent variable in the contrary way to how they affected life expectancy. Education is expected to be negatively correlated with cause of death since it is used as an indicator for smoking and less smoking should, on average and in theory, lead to less sensitivity or likelihood to suffer from respiratory and/or cardiopulmonary illnesses. Smoking and air pollution are not, however, the only causes of such illnesses. Income has a positive effect on cause-specific mortality, the opposite of the direction of its coefficient in the 2010 regression using life expectancy. The same is true for population change; population change is negatively correlated with cause of death, the opposite as to when using in correlation with life expectancy.

It is interesting, and this is true for the cause-specific mortality regression for 1990 and 2000 as well, that R^2 is very high. It seems improbable that these few factors should cover a large amount of the variation in the cause of death.

Fixed Effects Model:

Disclaimer: the correct way to include the different years in the fixed effects model would be to include a linear time trend in place of the year dummies that are included in the regression models below. Regressions run with a time trend variable instead of the year dummies were run but are not included in the results since they did not improve the result. The idea was to see if the incorrect use of the year dummies was what caused the unexpected direction of the effect of PM_{25} emissions in the model. The inclusion of the linear time trend did not, however, cause any changes in the direction of the effect of fine particle pollution nor did the change affect the size of the effect to a large degree.

The R^2 included in the tables of result is only the overall R^2. The model uses robust standard errors so adjusted R^2 is not calculated in the regression output.

Table 20: Summary Statistics for 2005-2011

	2005	2006	2007	2008	2009	2010	2011
Life expectancy at birth (e0), in years	80.748	80.938	81.165	81.312	81.169	81.709	81.811
Life expectancy at age 40 (e40), in years	41.739	41.97	42.15	42.287	42.097	42.637	42.777
Fine particle pollution (pm2.5), in thousands of tons	2299.047	978.005	1827.992	2989.621	979.422	1483.953	1831.325
Education, in percentage of population with at least high school diploma	0.529	0.475	0.448	0.505	0.467	0.442	0.487
Gross regional income, adjusted for inflation	116.754	101.752	93.788	103.636	97.706	101.979	106.539
Change in population from previous year, in percent	0.001	0.003	0.004	0.004	0.005	0.004	0.003
Cause of death (COD) for cardiopulmonary and lung diseases per 100,000 of population	543.654	529.063	515.786	513.685	502.225	481.788	479.638

Table 20: Summary Statistics for 2005-2011, source: various (for details see "Descriptive Statistics" sub-section above)

Table 21: Fixed Effects Model with Life Expectancy at Age 40 as the Dependent Variable

e40	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
$PM_{2.5}$	0.0000	0.0001*	0.0001*	0.0001	0.0001*	0.0001*	(0.0001)
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Education		-1.7347	-1.5959	-1.4773	-3.5925	-4.0541*	-3.7103*
		(1.4050)	(1.3673)	(1.3387)	(1.3516)	(1.0963)	(0.9081)
Income			-0.0009	-0.0003	0.0005	-0.0008	-0.0001
			(0.0027)	(0.0025)	(0.0025)	(0.0021)	(0.0036)
Population				8.4587*	0.0120	2.0304	3.9148
Change				(3.1819)	(3.8288)	(2.9148)	(3.5436)
Cause of Death					-0.0128*	-0.0116*	-0.0094*
					(0.0027)	(0.0018)	(0.0012)
2006	0.2967*	0.2278*	0.2299*	0.2211*	-0.0305	-0.0840	-0.0792
	(0.0597)	(0.0819)	(0.0831)	(0.0702)	(0.0833)	(0.0785)	(0.0882)
2007	0.4344*	0.3025*	0.2968*	0.2952*	-0.1857	-0.2932	-0.2163
	(0.0497)	(0.1137)	(0.1192)	(0.1174)	(0.1492)	(0.1414)	(0.1496)

2008	0.5135*	0.4570*	0.4455*	0.4377*	-0.0228	-0.0293	0.0540
	(0.0713)	(0.0872)	(0.0998)	(0.0986)	(0.1378)	(0.1147)	(0.1839)
2009	0.4228	0.3403	0.3401	0.3188	-0.2321	-0.2768	-0.1377
	(.2749)	(0.3263)	(0.3277)	(0.1340)	(0.4276)	(0.2105)	(0.1889)
2010	0.9387*	0.8021*	0.8062*	0.7942*	-0.1311	-0.1715	-0.0189
	(0.0597)	(0.1336)	(0.1346)	(0.1340)	(0.2072)	(0.1881)	(0.1322)
2011	1.0610*	0.9964*	0.9962*	0.9901*	-0.1121	0.1378	0.2811*
	(0.0731)	(0.0850)	(0.0853)	(0.0857)	(0.1890)	(0.1420)	(0.1247)
Constant	41.6251*	42.4994*	42.5126*	42.3934*	50.3357*	49.8588*	48.7058*
	(0.0586)	(0.7333)	(0.7471)	(0.7341)	(1.6904)	(1.3119)	(0.9045)
R^2	0.1813	0.1756	0.1774	0.2301	0.3231	0.3269	0.3424

Table 21: FE with Life Expectancy at Age 40

$PM_{2.5}$ is statistically significant and positive in model 2,3,5 and 8. The effect of the pollution on life expectancy is positive which is the opposite of what is expected. Other forces may be at large that causes this change or the model and/or data may be defective. In the cross-sectional models an explanation for the contrary effect of the fine particle pollution was that high pollution counties are also high life expectancy counties. Stockholm, Skåne and Västra Götaland counties have been among the highest polluting counties throughout the years but do also place at the top or above average for life expectancy, education and income. This may skew the results and make it seem like high pollution leads to high life expectancy since for high population density counties this is true. There are also low pollution counties such as Norrbotten that have lower levels of life expectancy despite the low pollution.

The effect of education is also the opposite of what is expected; the direction of the coefficient of education is negative in all models and the variable is statistically significant in model 6 and 7. Education may carry the effect of high air pollution since high education counties are often counties where the level of emissions is high (see the argument in the previous paragraph). Income is negative in all models but model 5 and not statistically significant in any. Income may also affect the pollution level if the increased income affected the production of goods that are produced by heavy emission industries (examples of such industries are the energy sector and steel production). If increased income affects the emissions level then it may be negatively correlated with life expectancy. Change of population is statistically significant in model 4 and the direction of its coefficient is positive in all models where included. This is what was expected in the theory. Cause of death is negative and statistically significant in all models. This effect is as expected and may account for the pollution level and could absorb the effect of the pollution. The intercept changes between the different years: 2005 is the base year and for some of the years the difference in intercept is statistically significant (see table)

Table 22: Fixed Effects Model with Life Expectancy at Birth as the Dependent Variable

e0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
$PM_{2.5}$	0.0000	0.0002	0.0001	0.0001	0.0001	0.0001*	0.0001
	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Education		-2.2657	-2.3013	-2.2174	-4.2934*	-4.4164*	-4.1287*
		(1.4937)	(1.555)	(1.5426)	(1.4899)	(1.3292)	(1.3977)
Income			0.0022	0.0008	0.0016	0.0049	0.0016
			(0.0031)	(0.0029)	(0.0034)	(0.0026)	(0.0004)
Population				9.0575*	0.5951	3.113	4.1757
Change				(3.8322)	(4.1702)	(2.7377)	(3.1260)
Cause of Death					-0.0123*	-0.0115*	-0.0099*
					(0.0027)	(0.0017)	(0.0010)
2006	0.3251*	0.1450	0.1445	0.1350	-0.1170	-0.0131	-0.1338
	(0.0743)	(0.0874)	(0.0893)	(0.0790)	(0.0963)	(0.0973)	(0.1019)
2007	0.4328*	0.2605*	0.2619*	0.2602*	-0.2216	-0.2671	-0.1655
	(0.0546)	(0.1123)	(0.1184)	(0.1190)	(0.1420)	(0.1605)	(0.1358)
2008	0.5410*	0.4669*	0.4699*	0.4615*	0.0459	0.0208	0.1353
	(0.0587)	(0.0761)	(0.0939)	(0.0933)	(0.1281)	(0.1131)	(0.1385)

2009	0.4657	0.3579	0.3580	0.3352	-0.2583	-0.1235	-0.1719
	(0.2841)	(0.3398)	(0.3413)	(0.3438)	(0.1966)	(0.2059)	(0.1527)
2010	0.9883*	0.8098*	0.8087*	0.7959*	-0.1311	-0.1235	0.0436
	(0.0639)	(0.1435)	(0.1445)	(0.1460)	(0.1695)	(0.2009)	(0.1830)
2011	1.0790*	0.9946*	0.9947*	0.9882*	0.1086	0.1950	0.3103*
	(0.0728)	(0.0870)	(0.0874)	(0.0901)	(0.1695)	(0.1329)	(0.0978)
Constant	80.6699*	81.8117*	81.8083*	81.6807*	83.6379*	88.9297*	88.0848*
	(0.0655)	(0.7726)	(0.7800)	(0.7730)	(81.5609)	(1.2729)	(0.8478)
R^2	0.1742	0.1726	0.1722	0.2276	0.3609	0.3693	0.3824

Table 22: FE with Life Expectancy at Birth

$PM_{2.5}$ is only statistically significant in model 6, but the direction of its coefficient is positive or zero in all which is the opposite of the expected. This mirrors the result in the regression output for the model above where life expectancy at age 40 is the dependent variable. The explanation of why this may be is the same. Education is negative, unexpected, in all and statistically significant in model 5, 6 and 7. This is also similar to the result above. Income is never statistically significant but is positive in all models where included. It is interesting that income has a positive effect on life expectancy at birth but not at age 40. Population change is positive but only statistically significant in model 4. According to theory, population change is expected to have a positive effect on life expectancy. The variable of population change can also be a confounding variable since it can be influenced by life expectancy.

Cause of death is statistically significant and negative, the expected direction of the effect, in all models where included. The cause-specific mortality is an intermediary between PM_{25} and life expectancy so the negative and statistically significant effect of cause of death may perhaps even out the positive effect of PM_{25} on life expectancy.

The intercept, or constant, for the different years differs from the base year, 2005, and, as can be seen in the table above, that difference is statistically significant in some models and for some years (noticeably not for the year 2009). R^2 decreases with the inclusion of education and income but increases with population change and cause of death. This is strange since R^2 should increase with added variables. It could be that the overall R^2, which the version included in the tables, decreases but that within and between R^2 increase.

Table 23: Fixed Effects Model with Cause of Death as the Dependent Variable

Cause of Death	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$PM_{2.5}$	-1.64x10^-6	0.0015	0.0009	0.0014	0.0015	0.0003
	(0.0017)	(0.0018)	(0.0021)	(0.0020)	(0.0017)	(0.0030)
Education		-139.1439*	-159.2462*	-165.5137*	-166.6073*	-134.7353*
		(58.0180)	(60.3310)	(56.3026)	(54.1632)	(58.9669)
Income			0.1050	0.0646	-0.0489	0.0325
			(0.1499)	(0.1452)	(0.1474)	(0.2125)
Population				-660.6862*	-622.6781*	-602.6726*
Change				(98.557)	(108.3008)	(141.7641)
2006	-14.5931*	-20.1176*	-20.3780*	-19.6862*	-19.8031*	-17.0216*
	(4.6769)	(5.2271)	(5.1968)	(4.4404)	(4.8350)	(5.8487)
2007	-27.8684*	-38.4520*	-37.7537*	-37.6278*	-37.1236*	-38.5903*
	(4.6961)	(6.6784)	(6.5838)	(4.4638)	(6.8856)	(6.7423)
2008	-29.9680*	-34.4984*	-33.0733*	-32.4654*	-32.778*	-28.1077*
	(6.1737)	(6.1081)	(6.3114)	(6.1903)	(5.4221)	(10.2743)
2009	-41.4312*	-48.0488*	-48.0174*	-46.3533*	-50.9187*	-45.7366*
	(5.3893)	(5.5597)	(5.5462)	(5.5071)	(5.3903)	(5.1583)

2010	-61.8675*	-72.8303*	-73.3400*	-72.4022*	-73.4476*	-65.8149*
	(5.1782)	(7.4874)	(7.5476)	(7.4372)	(7.2894)	(10.6275)
2011	-64.0165*	-69.1989*	-69.1727*	-68.7001*	-70.3054*	-62.2387*
	(6.0137)	(6.8814)	(6.8786)	(7.0183)	(5.6981)	(9.4290)
Constant	543.6576*	613.7786*	612.1502*	621.4651*	606.8303*	595.8673*
	(4.3736)	(30.0395)	(30.6773)	(30.1641)	(29.8004)	(40.2528)
R^2	0.0832	0.0719	0.0707	0.1523	0.1493	0.1531

Table 23: FE with Cause of Death

$PM_{2.5}$ is never statistically significant and the direction of its coefficient is positive in all models but model 1. The expected direction of the effect is positive since more pollution, theoretically, should lead to higher numbers of deaths due to diseases that may be exacerbated or caused by air pollution. In the cross-sectional regressions, PM_{25} often had a negative effect on cause-specific mortality, which is the opposite of what was expected from theory. Here the direction of the effect is in line with theory. Education is statistically significant and negative in all models, which is to be expected. It is, however, not the opposite of the direction that the coefficients of education are for life expectancy in the previous two fixed effects models. Education is used as an indicator of smoking habits and a higher level of education represents a population that smokes less. Income is never statistically significant and its coefficient is positive in all models but model 5. Income might have a positive effect on cause-specific mortality if an increase in income leads to an increase in emissions or other unidentified forces that are hazardous to health. Population change is consistently negative and statistically significant. This is also in line with theory and the opposite of the effect of the variable in the life expectancy fixed effects models above.

The yearly intercepts differ from the base year and in all models for all years the intercept is lower than in the base year of 2005 and it is statistically significant. R^2 decreases with education and income but increases with population change, which is curious since R^2 should increase with the inclusion of more variables, regardless if the variables are “good” or “bad”.

Discussion:

The hypothesis based on previous research, mainly on the United States, presented in the theory chapter of this essay was that air pollution in the form of $PM_{2.5}$ would be positively correlated with mortality (of e.g. cardiopulmonary and lung diseases) and thus negatively correlated with life expectancy. The theory also stated that because of the differences between the United States and Sweden in demography and geography, for example, the effect would be smaller in Sweden than in the United States.

The results in the previous chapter show that air pollution in the fixed effects regressions using panel data for the 21 Swedish counties for the years 2005-2011 has no or a positive effect on life expectancy, contrary to the hypothesis. In the cross-sectional, single year, OLS regressions, air pollution does have the expected negative effect on life expectancy for the years 2000 and 2010 but not for 1990. The direction of the effect of air pollution is the same regardless if life expectancy at birth or at age 40 is used though when using life expectancy at birth the coefficient for $PM_{2.5}$ is only statistically significant for 2010. The direction of the effect is the same in all variations of the different regressions, meaning that it does not matter if for example the variables are weighted by either the square root of the population density or the inverse of the county area or if no weights are used.

When regressing the air pollution level, $PM_{2.5}$, on life expectancy, either at age 40 or at birth, $PM_{2.5}$ is never statistically significant when used as the single explanatory variable. This is the variant of the regressions called "Model 1" in the results section. The main version of the regression model is the model where education, income, population change and cause of death are added as control variables. In this version, Model 5, $PM_{2.5}$ is statistically significant and negative for 2000 and 2010 using life expectancy at age 40.

A regression model using cause of death as the dependent variable is also run for all years. The expected direction of the main explanatory variable, $PM_{2.5}$, is in this model the opposite of those when using life expectancy. The effect of $PM_{2.5}$ on cause of death is thus expected to be positive. In contrast to the other models, $PM_{2.5}$ is never found to be

statistically significant as an independent variable explaining the variance in cause of death. The direction of the effect also differs between the years and different versions of the regression model.

Since air pollution is damaging to health through long-term exposure, looking at the level of air pollution in one year and its correlation with life expectancy in that same year does not show the entire picture. A lagged effects variable could be used to see if the PM_{25} levels in previous years affect the year in question. Regressions with the fine particle pollution levels from 1990 and 2000 were run with 2010's life expectancies and cause-specific mortality. These results are not included because no significant differences between those models and the ones using only the PM_{25} from 2010 were found.

To see the effect of the pollution level on life expectancy or cause of death, a fixed effects model using panel data for the years 2005 to 2011 is used. Contrary to the theory, however, the coefficient, though seldom statistically significant, of $PM_{2.5}$ is positively correlated or not correlated at all with life expectancy. In the case of cause of death the result is as expected, pollution is positively correlated with mortality (cause of death) except from when using it as the sole independent variable.

Throughout the results chapter, an explanation for why the effect of PM_{25} has a positive effect on life expectancy has been that the distribution of the pollution is higher in counties where the life expectancy also is high due to other factors such as the included education and income levels that are high in those counties and other unknown or unobserved factors. The opposite is sometimes also true since in some low pollution counties the life expectancy is also lower than average. Examples of high pollution counties with higher than average life expectancies have been Stockholm, Skåne and Västra Götaland. Examples of low pollution counties with lower than average life expectancies are Gotland, Västerbotten and Norrbotten.

Comparison: Cross-Sectional Data

The first part of the results chapter presents tables of results for the cross-sectional linear regressions for the years 1990, 2000 and 2010. It shows the estimates found using linear regression models with life expectancy at age 40, life expectancy at birth and cause of death as the dependent variables respectively. For this part of the discussion, the results of those regressions are compared. In the results chapter up to seven different models were presented which are made up of differing numbers of independent variables and different weights. In the tables below, two models are shown: Model 5 in which $PM_{2.5}$, education, income, population change and cause of death are the explanatory variables and Model 4 in which cause of death is as the dependent variable and not as an explanatory variable.. As explained in the method section of this essay, robust standard errors are used to correct any heteroskedasticity present. The use of robust standard errors also means that no adjusted R^2 are calculated. This means that R^2 cannot be used as an indicator of whether the addition of a variable is "good" or "bad".

Table 24: Comparison between Results for Regressions Using Life Expectancy at Age 40 as the Dependent Variable

	1990	2000	2010
e40	Model 5	Model 5	Model 5
$PM_{2.5}$	-0.0000 (0.0001)	-0.0003* (0.0001)	-0.0002* (0.0001)
Education	-7.3564* (3.2677)	4.2045 (4.5886)	0.3941 (4.1398)
Income	0.0075 (0.0095)	0.0025 (0.0111)	-0.0019 (0.0036)
Population Change	100.4353* (38.3117)	46.4702 (24.3296)	71.9459* (31.8458)
COD	-0.0033 (0.0017)	-0.0002 (0.0028)	-0.0012 (0.0022)
Constant	42.3256* (1.8778)	39.7585* (3.0957)	43.3563* (2.7602)
R^2	0.5471	0.5237	0.6238

Table 24: Cross-sectional Comparison, LE40

For life expectancy at age 40, fine particles emissions are negatively correlated with the dependent variable in year 2000 and 2010 but zero in 1990. No variable is consistently statistically significant in all three years nor is the magnitude of the coefficients the same, the standout being education for which the coefficient is much lower in 2010 than in the other two years. The direction of the coefficient is also negative in 1990 which is the opposite of what one might expect from the theory that higher educated people are more likely to belong to a higher socioeconomic class and are thus less likely to smoke which would in turn increase their life expectancy. Another variable, though never statistically significant, for which the coefficient changes sign, is income. The effect of income is positive in 1990 and 2000 but negative in 2010. The expected effect of income, according to the theory, was ambiguous so either direction may have been expected though the change of direction is more of a surprise.

Table 25: Comparison between Results for Regressions Using Life Expectancy at Birth as the Dependent Variable

	1990	2000	2010
e0	Model 5	Model 5	Model 5
$PM_{2.5}$	-0.0000 (0.0001)	-0.0002 (0.0001)	-0.0002* (0.0001)
Education	-6.8613 (3.9654)	0.9364 (7.0163)	2.4260 (5.2749)
Income	0.0067 (0.0117)	0.0087 (0.0155)	-0.0049 (0.0033)
Population Change	117.2864* (49.4709)	39.1331 (29.7684)	95.8356* (28.6453)
COD	-0.0027 (0.0026)	-0.0011 (0.0035)	-0.0007 (0.0021)
Constant	80.2661* (2.8450)	79.8866* (4.6589)	80.8506* (3.3899)
R^2	0.4809	0.3006	0.6516

Table 25: Cross-sectional Comparison, LE0

For life expectancy at birth, $PM_{2.5}$ is only statistically significant in 2010 but the direction of the coefficient is negative or zero in all years. Education and income are not

statistically significant in any year but the sign of their coefficients changes; the effect of education is negative in 1990 but positive, of different magnitudes, in 2000 and 2010 and the effect of income is positive in 1990 and 2000 but negative in 2010. Population change and cause of death follow the same pattern as with life expectancy at age 40.

Table 26: Comparison between Results for Regression Using Cause of Death as the Dependent Variable

	1990	2000	2010
COD	Model 4	Model 4	Model 4
$PM_{2.5}$	-0.0082 (0.0095)	- 0.0081 (0.0102)	-0.0026 (0.0067)
Education	-752.2535 (376.3739)	-1212.338* (258.1146)	-880.9936* (371.8678)
Income	-0.3855 (1.113)	1.5700* (0.6220)	0.3614 (0.3989)
Population Change	-11269.31* (4216.684)	-5804.752* (2328.034)	-7498.915* (2864.342)
Constant	1028.763* (41.7699)	958.7182* (80.6518)	902.2563* (170.3221)
R^2	0.6849	0.8540	0.8390

Table 26: Cross-sectional Comparison, COD

$PM_{2.5}$ is never statistically significant when used as an explanatory variable for cause of death but the coefficient is consistently negative which is the opposite direction of what is expected from the theory since a larger amount of pollution is expected to lead to a larger number of deaths from pollution related causes. The coefficient of education is negative, as expected, but only statistically significant in the two later years. The effect of income changes between the years, though as discussed before the expectation of the effect is ambiguous. Population change is consistently negative and statistically significant and the direction is the opposite of that when life expectancy is used as the explained and the effect is therefore as expected.

$PM_{2.5}$:

There is no surefire way to use the $PM_{2.5}$ emission data that is reported in tons per year and calculate estimates for the averages of pollution levels per day or hour in $\mu g/m^3$, the measure used in previous analysis. Even if factors such as pollution being absorbed by for example vegetation and water is ignored, one would need to know the size in cubic meters of a county to be able to calculate an estimate for the average level of fine particles in the air. Since there is no proven scientific way of transforming the data, the emission levels are left in their original states in the statistical analysis. Before settling on using $PM_{2.5}$ in its original form (thousands of tons per year) several different forms were used such as per square kilometer or per capita, none of which gave any "better" results.

There are official Swedish data for fine particle levels in the air in the form of $\mu g/m^3$ for the later years used in this analysis. The choice not to use this data was made because it is sporadic in the years and areas of availability as well as whether the figures regarded hourly or daily yearly averages (Naturvårdsverket). The data available also differ between in what kind of area the measurements were taken (Ibid.). Both data for rural and urban areas are available, the years and hourly/daily averages of which, however, differ. Because of the inconsistencies of the available data the number of observations in the analysis would be lower which could reduce the accuracy of the models. In addition, since the figures are for smaller geographical areas, some of the control variables would be unavailable (e.g. regional income).

Time Period:

Since air pollution affects morbidity and mortality slowly, a longer time frame would be preferable. Previous research has shown that short run changes in pollution have no effect on life expectancy (Chay et. al., 2003; Beverland et. al., 2012). The fixed effects model using panel data for the years 2005 to 2011 is an attempt to see if there is an effect of particle pollution on life expectancy in the long run. In the results for this model, however, the effect of the pollution is positively correlated with life expectancy. This is the opposite of the expected direction of the effect. It may be that the negative effect of the pollution is "swallowed" by the mediator variable cause of death. Cause of death, the independent variable, is only statistically significant in the fixed effects

model, so this may indicate that the effect of cause of death can only be seen when using panel data.

In addition to the fixed effects model, an OLS regression model for 2010 was run that included the PM_{25} levels for 1990 and 2010 to see if there were any lagged effects. The result was not different enough to be significant and the result is therefore not included in the essay. There should, logically and according to previous research, be a lagged effect of the pollution level since long term exposure is needed for the particulate matter to be hazardous "enough" to cause damage. Any better alternative than the fixed effects model, where the time span used is still on the short side, was not found.

Geographical Areas:

The econometric analysis may have benefitted from the use of municipalities instead of counties. If municipalities had been used the number of observations would have increased greatly (Sweden has 21 counties and 290 municipalities). A larger number of observations could have increased the accuracy of the outcomes of the regression models. Previous research on air pollution in Sweden shows that the size of the geographical area can be of great importance and that the smaller the areas the better (Stroh et. al., 2005). The decision not to use municipalities is two-fold: one, gross regional product is not available for municipalities and two, the calculation of life expectancy for each municipality for each year would have been too time consuming.

The level of pollution may also differ greatly within a county, which could distort the analysis. An example of this could be Norrbotten County where the area is vast but the pollution is concentrated in a few municipalities. Luleå, for one, has a large steel industry that might account for its higher levels of particle pollution. Kiruna and Gällivare, for further example, have large mining industries that may affect the environment in the same way. The population of Norrbotten county is, however concentrated in the same places as the higher pollution level. In line with the theory, the life expectancy in Norrbotten is low. The observations are, on the other hand, outliers when compared to the other counties. A contributing factor to the low life expectancy can be that the use of tobacco products is higher in the north of Sweden (scb.se, 2015).

Stockholm, Scania and Västra Götaland county, the counties that are home to the three large metropolitan areas (Stockholm, Malmö and Gothenburg) have statistics that are contrary to the theory that higher air pollution leads to lower life expectancy. The pollution level in these counties is high yet so is the life expectancy. The high level of emissions may be due to the large industries in the areas as well as heavy traffic. The higher level of education and the lower consumption of tobacco products could be seen as two factors that are positively correlated with the high(er) life expectancy.

Differences Sweden and United States:

It is not possible to outright compare this analysis with the analysis by for example Correia et. al. (2013) since that study looks at the effect of different levels of regulation, it uses different data for PM_{25} pollution levels (concentration in air versus emissions in tons) and the number of observations are much larger in that study. Their model also uses the changes in the variables between two years and not a fixed effects model or cross-sectional models with only one year.

If one disregards the size of the country and the population, there are still disparities between the two countries that may lead to the difference in result from what was expected from the theory. As mentioned before, Sweden has a consistently higher life expectancy and lower level of emission (not concentration in air) of the fine particle pollution. This may affect how large an effect a change in PM_{25} could have on life expectancy.

Conclusion:

This analysis attempts to see if the fine particle pollution level in Sweden has a negative effect on life expectancy. The outcomes from the statistical models used in the analysis do not, however, show much indication that the pollution level has a significant effect on the life expectancy nor on cause-specific mortality. A reason for this may be that the time frame is too short to show any damaging effect of the air pollutants on health since it is caused by long-term exposure. Another reason for the lack of results may be that the areas are too large: the statistical analysis could have benefitted from the use of the 290 municipalities in place of the 21 counties. A third reason could be the concentration and composition of people in the metropolitan areas and that the life expectancies in counties with metropolitan areas have high life expectancies because of other factors despite having high levels of pollution.

The lack of any "good" results can also be because the quantity of emission of fine particle pollution is used and not the concentration of the pollutants in the air as has been done in previous analyses. It might be that the data on the level of emissions is not a good enough alternative to be able to discern any link between PM_{25} pollution and life expectancy or PM_{25} and cause-specific mortality. This might have also enabled an easier comparison between Sweden and the United States something that was difficult to do using this data.

In conclusion, if further research would be pursued on the effect of fine particle pollution on the life expectancy or mortality in Sweden, the use of smaller geographical areas and a longer time frame could improve the analysis and might lead to more conclusive results. If available, PM_{25} concentrations in the air in place of quantities of emission might bring more useful results.

References:

Chay, Kenneth, Dobkin, Carlos and Greenstone, Michael (2003). "The Clean Air Act of 1970 And Adult Mortality", The Journal of Risk and Uncertainty vo. 27 no . 3, 279-300

Correia, Andrew W., Pope, C. Arden, Dockery, Douglas W., Wang, Yun, Ezzati, Majid and Dominici, Francesca (2013), "Effects of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 Counties for the Period from 2000 to 2007", Epidemiology, vol. 24, no. 1, 23-31.

Douglas W. Dockery, C. Arden Pope III, Xiping Xu, John D. Spengler, James H. Ware, Martha E. Fay, Benjamin G. Ferris and Frank E. Speizer (1993). "An Association between Air Pollution and Mortality in Six US Cities", The New England Journal of Medicine, vol. 329, no. 24, 1753-1759.

Iain J. Beverland, Geoffrey R. Cohen, Mathew R. Heal, Melanie Carder, Christina Yap, Chris Robertson, Carole L. Hart and Raymond M. Agius (2012). "A Comparison of Short-term and Long-term Air Pollution Exposure Associations with Mortality in Two Cohorts in Scotland", Environmental Health Perspectives vo. 120 no. 9, 1280-1285

Lehmijoki, Ulla and Rovenskaya, Elena (2009). "Air Pollution Mortality in Denmark, Finland and Sweden", Finnish Yearbook of Population Research, 97-105.

Liu, Ben-Chieh (1979). "The Cost of Air Quality Deterioration and Benefits of Air Pollution Control: Estimate of Mortality Costs of Two Pollutants in 40 US Metropolitan Areas", American Journal of Economics and Sociology vo. 38 no. 2, 187-195

Mariani, Fabio, Pérez-Barahona, Agustín and Raffin, Natacha (2010). "Life Expectancy and the Environment", Journal of Economic Dynamics & Control, no. 34, 798-815.

Meister, Kadri, Johansson, Christer and Forsberg, Bertil (2012)."Estimated Short-Term Effects of Coarse Particles on Daily Mortality in Stockholm, Sweden", Environmental Health Perspectives, vol. 120, no. 3, 431-436.

Pascal, M., Corso, M., Chanel, O., Declercq, C., Badaloni, C., Cesaroni, G., Henschel, S., Meister, K., Haluza, D., Martin-Olmedo, P. and Medina, S. (2013), "Assesing the Public Health Impacts of Urban Pollution in 25 European Cities: Results of the Aphekom Project", Science of Total Environment no. 449, 390-400.

Pope CA 3rd, Thun MJ, Namboodiri MM, et al. (1995) Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. American Journal of Respiratory Critical Care Medicine, vol. 151,669–674.

Samet, Jonathan M. (2013), "Air Pollution and Life Expectancy: Does Cleaner Air Lengthen Life?", Epidemiology, vol. 24, no. 1, 32-34.

Stroh, Emilie, Oudin, Anna, Gustafsson, Susanna, Pilesjö, Petter, Harrie, Lars, Strömberg, Ulf and Jakobsson, Kristina (2005), "Are Associations Between Socio-

economic Characteristics and Exposure to Air Pollution a Question of Study Area Size? An Example from Scania, Sweden", International Journal of Health Geographics, vol. 4, no. 30.

Wolff, Hendrik and Perry, Lisa (2010), "Trends in Clean Air Legislation in Europe: Particulate Matter and Low Emission Zones", University of Washington.

Web Sources:
US Environmental Protection Agency, Region 4:
http://www.epa.gov/region4/sesd/pm25/p2.html (last accessed: 5/1/15)
World Health Organization Air Quality Guidelines:
http://www.who.int/mediacentre/factsheets/fs313/en/ (last accessed: 5/1/15)

Data:
Statistiska Centralbyrån, www.scb.se
Regional Gross Product (last accessed: 5/30/14)
Cause of Death, 1990 (last accessed: 6/22/14)
Consumer Price Index (last accessed: 5/30/14)
Deaths (last accessed: 3/27/14)
Education (last accessed: 5/19/14)
Land Area (last accessed: 5/31/14)
Population (last accessed: 3/26/14)
Population Density (last accessed: 4/2/14)
Household expenditure 2000 (last accessed: 5/1/15)
Naturvårdsverket, www.naturvardsverket.se
$PM_{2.5}$ (last accessed: 4/2/14)
Socialstyrelsen, www.socialstyrelsen.se
Cause of Death, 2000-2011 (last accessed: 4/2/14)
Organisation for Economic Co-operation and Development, www.oecd.org
Life Expectancy (last accessed: 8/8/14)
$PM_{2.5}$ (last accessed: 8/1/14)

Appendix

n			1990 nPx	nDx	nMx	n nax	nqx	npx	lx	ndx	nLx	Tx	ex
1	0		**25173**	137	0.0054	0.1	0.0054	0.9946	100000	541.5812	99512.5769	7733881.673	77.3388
4	**1-4 yrs**		85644	49	0.0006	0.5	0.0023	0.9977	99458.41882	227.3547	397378.9658	7634369.096	76.7594
5	**5-9 yrs**		88714	21	0.0002	0.5	0.0012	0.9988	99231.06408	117.3783	495861.8746	7236990.131	72.9307
5	**10-14 yrs**		87862	13	0.0001	0.5	0.0007	0.9993	99113.68577	73.2968	495385.1867	6741128.256	68.0141
5	**15-19 yrs**		100324	39	0.0004	0.5	0.0019	0.9981	99040.38893	192.3181	494721.1493	6245743.069	63.0626
5	**20-24 yrs**		118513	81	0.0007	0.5	0.0034	0.9966	98848.07079	337.2219	493397.2991	5751021.92	58.1804
5	**25-29 yrs**		139820	110	0.0008	0.5	0.0039	0.9961	98510.84887	386.7445	491587.3832	5257624.621	53.3710
5	**30-34 yrs**		126599	117	0.0009	0.5	0.0046	0.9954	98124.1044	452.3755	489489.5833	4766037.238	48.5715
5	**35-39 yrs**		119336	152	0.0013	0.5	0.0063	0.9937	97671.72893	620.0551	486808.5070	4276547.654	43.7849
5	**40-44 yrs**		131593	249	0.0019	0.5	0.0094	0.9906	97051.67385	913.8818	482973.6649	3789739.147	39.0487
5	**45-49 yrs**		126657	340	0.0027	0.5	0.0133	0.9867	96137.79209	1281.7669	477484.5433	3306765.483	34.3961
5	**50-54 yrs**		88767	402	0.0045	0.5	0.0224	0.9776	94856.02524	2123.8316	468970.5472	2829280.939	29.8271
5	**55-59 yrs**		74059	585	0.0079	0.5	0.0387	0.9613	92732.19363	3591.5821	454682.0130	2360310.392	25.4530
5	**60-64 yrs**		72602	803	0.0111	0.5	0.0538	0.9462	89140.61155	4796.9703	433710.6320	1905628.379	21.3778
5	**65-69 yrs**		77040	1306	0.0170	0.5	0.0813	0.9187	84343.64125	6858.4021	404572.2011	1471917.747	17.4514
5	**70-74 yrs**		66326	1877	0.0283	0.5	0.1321	0.8679	77485.23919	10239.5710	361827.2686	1067345.546	13.7748
5	**75-79 yrs**		52425	2550	0.0486	0.5	0.2168	0.7832	67245.66824	14581.3311	299775.0134	705518.2774	10.4917
5	**80-84 yrs**		36160	3030	0.0838	0.5	0.3464	0.6536	52664.33711	18243.1624	217713.7796	405743.264	7.7043
5	**85-89 yrs**		17305	2455	0.1419	0.5	0.5236	0.4764	34421.17472	18023.6715	127046.6948	188029.4844	5.4626
5	**90-94 yrs**		5565	1369	0.2460	0.5	0.7616	0.2384	16397.5032	12488.5574	50766.1226	60982.78961	3.7190
5	**95-99 yrs**		1091	417	0.3822	0.5	0.9773	0.0227	3908.945825	3820.0853	9994.5158	10216.66705	2.6137
5	**100+ yrs**		94	57	0.6064	0.5	1.0000	0.0000	88.86049801	88.8605	222.1512	222.151245	2.5000

where n is number of yrs in age group; nPx is total population in age group; nDx is total deaths in age group; nMx age-specific death rate; nax is the average population who lived (nax infants may not be the same as the one used by Statistics Sweden); nqx is the probability of dying; npx is the probability of living; lx is the hypothetical population alive at the start of the year; ndx is the hypothetical number of deaths; nLx number of people who lived through the interval; Tx is the total number of people alive at the end of the period; ex is the life expectancy