**European Journal of Environment and Public Health**
2023, 7(2), em0130
e-ISSN: 2542-4904
https://www.ejeph.com

MODESTUM

Research Article

OPEN ACCESS

# The impact of smoking, overweight, and fine particulate matter air pollution on life expectancy: Estimations with county-level matched data for Germany

Aloys Leo Prinz [1*], David Johannes Richter [1]

[1] Department of Business and Economics, Institute of Public Economics, University of Muenster, Muenster, GERMANY
*Corresponding Author: aloys.prinz@wiwi.uni-muenster.de

**Citation:** Prinz AL, Richter DJ. The impact of smoking, overweight, and fine particulate matter air pollution on life expectancy: Estimations with county-level matched data for Germany. EUR J ENV PUBLIC HLT. 2023;7(2):em0130. https://doi.org/10.29333/ejeph/12681

**ARTICLE INFO**

Received: 25 May 2022

Accepted: 29 Nov. 2022

**ABSTRACT**

Smoking and overweight are well-known factors that shorten life expectancy. While these factors are seemingly controllable by the individual, risks from fine particulate matter pollution are less so. In this paper, these risks are studied with novel micro data at the county level for Germany and for the years 1999 to 2017. A matching approach is used to control for relevant differences between the counties. Unexpectedly, fine particulate matter pollution is not found to have a direct effect on life expectancy, when controlling for relevant covariates with the matching estimation model. In contrast, it had just such a negative effect in the robustness check and extensions with an OLS model. These additional OLS estimations provide evidence of a moderating effect of particulate matter pollution on the effects of smoking and overweight with respect to life expectancy.

**Keywords:** smoking, overweight, air pollution, life expectancy, matched data, Germany

## INTRODUCTION

It is well known that smoking and overweight, but also air pollution, are global risk factors to life expectancy [1]. There are important differences between the risks themselves, as well as the strength of the life expectancy effects, as the abovementioned risk factors study demonstrates.

Among the various risk factors, two groups of factors are salient: behavioral risks on the one hand, environmental and occupational risks on the other [1]. Drug use, alcohol consumption, smoking, dietary behavior, etc. are behavioral risks that are based on individual choice. Therefore, these risk factor can be dubbed lifestyle factors [2], at least for high-income countries. Air pollution, climate, water quality and others are environmental risks based on location. In comparison to behavioral risks, environmental (and occupational) risks are location-dependent and as such, less dependent on individual choice, although some individual options also exist in this respect. Not surprisingly, personal decisions [3] and location [4] are found to be decisive risks to life expectancy.

Since personal and locational variables are relevant and significant factors with respect to life expectancy, the objective of this study is to analyze empirically the impact of smoking, overweight and fine particulate matter on life expectancy with county-level micro data from Germany. This study has two novel features. The first is the usage of unpublished data from the German Federal Office of Statistics (Statistisches Bundesamt) on smoking and overweight, in combination with locational data on fine particulate matter pollution provided by the German Federal Environment Agency (Umweltbundesamt). The second one is the application of matching methods to control for covariates at the local level.

The first main result of this paper is that we do not find a statistically direct effect between fine particulate matter and life expectancy with the applied matching method, in contrast to [5] with multiple regression analysis. The second main result is, however, that fine particulate matter has a statistically significant moderating effect on the relationship between smoking/overweight and life expectancy. Note that the effects mentioned here are directed correlations and not causal effects.

The structure of the remaining paper is, as follows. In the next section, lifestyle risks of smoking and overweight, as well as fine particulate matter air pollution, are considered as indirect factors that are related to a number of diseases leading to certain medical causes of death. Next, we present the data base, descriptive statics of the variables, and the applied estimation method. We then present the empirical results. After that, the results are checked with a robustness test. The results are then discussed. And finally, conclusions of the study are presented.

# HEALTH BEHAVIOR, LOCAL ENVIRONMENT, AND HEALTH

It was presented a model of complex networks with subnetworks that are interconnected with each other and that play a major role in human diseases [6]. The entire network encompasses three layers [6]: the most basic layer, the *cellular network* (consisting of the metabolic network itself, the regulatory network, as well as protein-protein interactions), a middle layer with the *disease network* and a top layer with *social network* (social links, family ties and physical proximity). According to [6], the interactions within and between these networks are important to "… quantify the complex interlinked factors that may contribute to individual diseases." The network of the disease genome and the respective disease phenome, the so-called "diseasome", is demonstrated in [7], alongside the human disease network. However, even at the most basic cellular level, environmental perturbations interact with genome-transcriptome-proteome in pathogenesis [8].

The diverse systems involved in the development and causation of diseases, and the health status of an individual can also be classified into the genetic background, the person's biology which interacts with the individual's social and physical environment, and psycho-social factors such as personality, lifestyles, as well as attitudes and beliefs [2].

Several papers are relevant for this study. Besides the global analysis in [1], there is a recent study on individual health-related risk factors in Germany [9]. The method applied is that of meta-analyses of prospective studies to determine the relative effect of the included risk factors on life expectancy. The included risk factors are smoking, physical activity, sedentary behavior, alcohol consumption, overweight and obesity, sleep, coffee consumption, diet, social network, and social participation [9]. The results are losses of life expectancy those exposed to the particular risk, and average losses of life expectancy from risk factor exposure. The largest losses of *average* life expectancy are from smoking and overweight/obesity. The loss from smoking for men is 2.66 (95% CI 2.50-2.83), for women 1.53 (95% CI 1.33-1.74) years, and from overweight/obesity for men 1.65 (95% CI 1.52-1.80) and for women 1.37 (95% CI 1.17-1.59) years [9]. The individual risks in comparison to the respective healthiest category are highest for current smokers with a loss of 6.85 (6.78-6.92) years for men and 5.86 (5.81-5.93) years for men. The second highest individual risk in comparison to the respective healthiest category is also obesity, with a loss of 5.01 (4.97-5.05) years for men and 4.34 (4.33-4.40) years for women [9].

The social significance of status on health and life expectance, in Germany is documented by [10], with data from the Socio-Economic Panel for 2002 to 2016. People in the lowest income bracket have a higher mortality risk and a shorter life expectancy at the age of 65 years. Moreover, women, as well as men with low social status suffer from poorer health in general and from certain diseases such as diabetes and depression [11]. However, low status women and men are considerably less active in sports and are more often overweight/obese [11]. These results are descriptive statistics and not intended to imply causality. Nevertheless, there is evidence from the USA that a causal relationship is likely to exist [12]. These findings support the inclusion of income and education (as proxies for social status) in the analysis in this paper.

The regional distribution of years-of-life-lost (YLL), as well as life expectancy, is not equally distributed over Germany. It was shown that age standardized-YLL are, for instance, lower in the German states of Baden-Wuerttemberg and Bavaria [13]. However, they also show that there are even differences in the regions within German states. The latter result is supported by data on life expectancy at birth. This was analyzed at the level of all 402 German counties [14]. In effect, there are considerable differences of life expectancy among counties that spread between 75.8 and 81.2 years for men, as well as 81.8 and 85.7 years for women [14]. These regional results for Germany confirm the inclusion of the place of residence on the county level in the present empirical investigation.

As shown empirically by [15], there is an association between fine particulate matter air pollution, measured by $PM_{2.5}$ and $PM_{10}$ (i.e., particulate matter with a diameter of 2.5 and 10 µm, 1 µm=$10^{-6}$ m) and life expectancy in Germany at the county level. However, the effect is not distributed evenly over Germany, but more concentrated in the western part of Germany. Moreover, COVID-19 related fatalities are also associated with particulate matter air pollution in Germany, as found by [5]. These results exhibit the relevance of particulate air pollution for the determination of German district-level life expectancy.

With these empirical results, we develop our estimation model in the following. However, before starting, two issues with empirical studies concerning the determination of health effects are worth mentioning. The first can be called the *equifinality* [16] of (environmental and other) risk factors and human behavior. This means that a number of different factors contribute to the absolute or the relative risk of a certain disease [17]. A case in point is smoking, but also overweight, concerning, e.g., a disease like cancer. Moreover, also genetic factors play a major role in almost all diseases. This makes it difficult to study environmental and behavioral effects with respect to diseases [17, 18]. Instead, these effects can be studied with respect to life expectancy. In this respect, equifinality means that one can survive several diseases, but inevitably (of course) one of later diseases will be terminal. The crucial question is which factors contribute to a shorter or longer life expectancy. In this respect, for instance, even socio-economic status can have effects not only on health behavior, but also directly on health status. Socio-economic status is partly determined by individual choices, but also by the social and physical environment, and chance.

The second issue is the *multifunctionality* of (health-related) behavior [19]. Health behavior and lifestyles are not intended to be risks to life. They enhance the enjoyment of life. In this respect, diseases may be considered an accident or simply a cost. Economically, multifunctionality implies choice-related trade-offs. The enjoyment of lifestyles comes with a price tag that enforces choice in the form of individual cost-benefit analyses. Even if it is denied that one chooses a certain lifestyle with its consequences, one could decide otherwise. Accordingly, choices are inherent ingredients of life, whether they are conscious decisions or unconscious ones.

Smoking and lifestyles that lead to overweight (e.g., diet, level of physical activity, etc.) are chosen, even if they are influenced by genetics, epigenetics, age, gender, and status. These choices have costs in the form of negative health effects.

Even the place of residence is in effect chosen. Of course, the place of birth and of adolescence are determined by other people and circumstances. Nevertheless, adults have a degree of freedom to choose their place of residence, although external influences are seemingly stronger than for smoking and overweight/obesity.

These considerations lead to smoking and bodyweight-related behavior (or lifestyles). In the study presented here, these behavioral factors are the relevant elements of health behavior that are assumed to have an effect of health status (see e.g., [3] for empirical results). Because of the long-term nature of the effect, health-status effects are defined as those on life expectancy.

The research questions of this paper are, as follows:

1. What are the directed correlations between smoking and overweight on life expectancy in Germany by controlling for socio-economic status (measured by income and education) and environmental pollution (measured by particulate matter concentration on the county level)?
2. Besides a negative direct correlation with life expectancy, does particulate matter pollution have a moderator effect on the relation between smoking/overweight and life expectancy?

To the best of our knowledge, there is no theory or mechanism that explains such a moderating effect of particulate matter air pollution on the relationship between smoking/overweight and life expectancy. However, a potential mechanism could work, as follows: Suppose that there are direct or first-order negative effects of smoking/overweight and particulate matter pollution on life expectancy. As a second-order effect could be generated by particulate matter pollution as its first-order effects concern the human respiratory system [20-24]. This is also the case with smoking (although this is not the only effect), but cigarettes contain and emit also particulate matter [25]. By contrast, overweight seems to have a more general effect on the functioning of the human body and, therefore, health [26]. Since the human respiratory system is already damaged directly by the different ingredients of cigarettes and particulate matter, the second-order effect of outside particulate matter pollution might be smaller than the first-order effect of smoking. Nevertheless, it is not possible to say theoretically whether the combined effect of particulate matter *and* smoking is positive, negative, or even neutral. This can only be determined empirically.

This may be different for overweight because it has no direct relationship with particulate matter inhalation. A potential second-order effect of particulate matter could be that higher outdoor concentration may enhance the first-order effect of overweight on life expectancy. However, also in this case, it cannot be taken for granted that the effect is positive. The reason is that the combined effect captures the change of the already determined direct effect between overweight (or smoking) and life expectancy due to the combined effect.

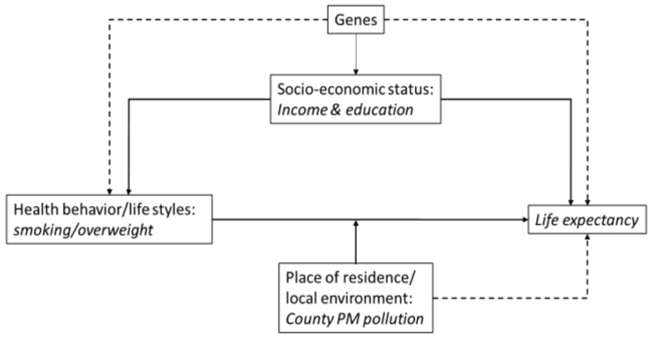


**Figure 1.** County particulate matter air pollution as a moderator variable between smoking/overweight and life expectancy (Source: Own depiction)

**Figure 1** shows diagrammatically our research questions. As is already well-known, as shown in section on the relevant literature above, there is a direct effect of smoking (see in addition [27, 28]) and bodyweight-related behavior, overweight, on life expectancy [26]. Furthermore, it is also known that socio-economic status has a direct effect on both lifestyles and life expectancy [12, 29]. The influence of place of residence is represented by county-level particulate matter air pollution. The latter may have a direct effect on life expectancy, as shown in [5]. Furthermore, as explained above, particulate matter pollution may also have a moderating effect on the relation of smoking/overweight and life expectancy.

For the following empirical investigation, no individual data are available. Nonetheless, county data are present and can be used. In this respect, the effects defined by the above graph are not 'causal' individually. Therefore, the study presented is an observational one and it is not claimed here that the effects found and presented below are 'causal' (see [30] for causal modeling in environmental health, as well as [17]). The effects should be understood as directed correlations.

## SMOKING, OBESITY, FINE PARTICULATE MATTER AIR POLLUTION, AND LIFE EXPECTANCY IN GERMAN COUNTIES

### Data

The data used in this study are from two official German sources, namely the Federal Office of Statistics and the Federal Environmental Agency. The data on fine particulate matter air pollution, $PM_{2.5}$ and $PM_{10}$, are publicly available on the homepage of the Federal Environment Agency [31]. $PM_{10}$-data are reported since 2002, $PM_{2.5}$-data since 2010. The data used here are aggregated to county-level. This is the most extensive data on fine particulate matter air pollution in Germany.

By contrast, the data on smoking and overweight people in Germany are collected by the German Micro Census, but not publicly available. The Micro Census is the most extensive annual household survey in Germany, carried out by the German statistics office [32]. About 1% of the population–810,000 people in 370,000 households–are asked about their living and working conditions [32].

**Table 1.** Descriptive statistics without data on particulate matter pollution

| | n | Mean [95% CI] | SD | Min. | Pctl (25) | Pctl (75) | Max. |
|---|---|---|---|---|---|---|---|
| Life expectancy [*years*] | 2,197 | 79.638 [79.580; 79.696] | 1.383 | 75.330 | 78.660 | 80.610 | 83.600 |
| Smokers [*share*] | 2,197 | 0.241 [0.239; 0.242] | 0.038 | 0.090 | 0.214 | 0.267 | 0.415 |
| Overweight [*share*] | 2,197 | 0.439 [0.437; 0.441] | 0.054 | 0.256 | 0.404 | 0.474 | 0.625 |
| High school diploma [*percentage*] | 2,197 | 27.257 [26.831; 27.683] | 10.180 | 0 | 20.100 | 33.600 | 64.000 |
| Unemployment [*per 1,000 persons**] | 2,197 | 61.159 [59.942; 62.376] | 29.081 | 11.100 | 40.600 | 75.400 | 187.200 |
| Income [*1,000 Euros*] | 2,197 | 18,567.280 [18,436.53; 18,698.03] | 3,125.119 | 11,490 | 16,317 | 20,524 | 35,587 |
| Urban [*1=Urban*] | 2,197 | 0.514 [0.493; 0.535] | 0.500 | 0 | 0 | 1 | 1 |
| East [*1=East Germany*] | 2,197 | 0.160 [0.145; 0.176] | 0.367 | 0 | 0 | 0 | 1 |

Note. n: Number of observations; SD: Standard deviation; Min: Minimum value; Max: Maximum value; Pctl (25): 25 percent percentile; Ptcl (75): 75 percent percentile; & *: Unemployment per 1,000 persons of working age

**Table 2.** Descriptive statistics including data on $PM_{10}$-background pollution

| | n | Mean [95% CI] | SD | Min. | Pctl (25) | Pctl (75) | Max. |
|---|---|---|---|---|---|---|---|
| Life expectancy [*years*] | 826 | 79.822 [79.736; 79.909] | 1.263 | 75.330 | 78.953 | 80.627 | 83.600 |
| Smokers [*share*] | 826 | 0.246 [0.243; 0.248] | 0.036 | 0.143 | 0.219 | 0.271 | 0.360 |
| Overweight [*share*] | 826 | 0.445 [0.441; 0.449] | 0.059 | 0.256 | 0.405 | 0.484 | 0.609 |
| $PM_{10}$-background [$\mu g/m^3$] | 826 | 20.152 [19.796; 20.508] | 5.219 | 3 | 17 | 23 | 41 |
| High school diploma [*percentage*] | 826 | 30.730 [30.021; 31.439] | 10.381 | 6.800 | 23.000 | 37.575 | 64.200 |
| Unemployment [*per 1,000 persons**] | 826 | 69.162 [67.106; 71.218] | 30.103 | 17.700 | 46.825 | 86.275 | 186.000 |
| Income [*1,000 Euros*] | 826 | 18,700.280 [18,492.19; 18,908.36] | 3,046.831 | 13,127 | 16,666 | 20,159.2 | 35,587 |
| Urban [*1=Urban*] | 826 | 0.580 [0.546; 0.614] | 0.494 | 0 | 0 | 1 | 1 |
| East [*1=East Germany*] | 826 | 0.260 [0.230; 0.290] | 0.439 | 0 | 0 | 1 | 1 |

Data referring to overweight are available for the years 1999, 2003, 2005, 2009, 2013, and 2017, referring to smoking behavior additionally for the year 1995. The data used in this paper are form a special analysis of the Micro Census, provided by the statistics office [33]. In addition, data on the average life expectancy of newborns, disposable income, education (share of school leavers with a high school certificate) and urban counties are publicly available from the statistics office [34]. Furthermore, we included a dummy variable for East Germany.

## Descriptive Statistics of All Variables

As described above, data for PM, overweight, and smoking behavior are available for different time ranges. Considering the above variables and eliminating incomplete data because of missing values, final datasets are resulted. Such an extensive overview of the descriptive statistic is necessary, since the number of observations, as well as the composition of the data, changes when taking account particulate matter concentration especially. Note that the share of overweight people is defined as the share of people with a body mass index (BMI) greater or equal $25\frac{\text{kg}}{\text{m}^2}$. Moreover, note also that all empirical results in this paper are based on data from the abovementioned sources [32-34].

As **Table 1** shows, there are 2,197 observations, referring to the period 1999 to 2017, when data for fine particulate matter pollution is not considered. If data on particulate matter pollution with a diameter of 10 µm and less, $PM_{10}$, is included, 826 observations on background pollution for the period from 2003 to 2017 are available. By contrast, there are 423 observations for traffic area pollution. For particulate matter concentration with a diameter of less than 2.5 µm, $PM_{2.5}$, 143 observations for background pollution and 70 observations for traffic area pollution are available. The most comprehensive observations are described statistically in **Table 1**. The most interesting values are the share of smokers in the population over the period 1999 to 2017 (24.1%) and the share of overweight people (43.9%). With an average urbanization of 51.4% and 16.0% of East German observation in the data, Germany seems adequately represented.

Adding data on background $PM_{10}$ pollution not only reduces the number of observations considerably, but also changes the composition of the observations as in **Table 2**, where the share of East German observations increases to 26%. Moreover, the urbanization index is 58%. Nevertheless, the shares of smokers and overweight people remain quite stable. Also, life expectancy and share of smokers increase significantly, as is evident by comparing 95% CIs.

Similarly, accounting for $PM_{10}$-traffic data in **Table 3** increases the urbanization rate of the data to 74%. This was to be expected since traffic-related recording stations are mainly installed in cities. In addition, East Germany seems somewhat overrepresented with 24.8% of the observations. Comparing the confidence intervals, life expectancy and the share of smokers do not deviate significantly from the values in **Table 2**, but from the values in **Table 1**. The share of smokers is significantly lower in **Table 3** than in **Table 2**.

Data on $PM_{2.5}$ are still relatively scarce in Germany. Moreover, as shown in **Table 4** for $PM_{2.5}$-background data, East Germany (30.8% of the observations) is better covered than West Germany. Urbanization is also somewhat higher (59.4%) than in the most comprehensive **Table 1**, but considerably lower than in **Table 3**. However, **Table 2** is more relevant for a comparison, as it contains data on $PM_{10}$-background pollution, and **Table 4** shows data for $PM_{2.5}$-background pollution.

In this respect, urbanization is similar (58% in **Table 2** and 59.4 in **Table 4**), whereas the share of East German observations is at 30.8% higher, in **Table 4** than in **Table 2** (26%). The shares of smokers (24.6% in **Table 2** versus 23.6% in **Table 4**) and overweight people (44.5% in **Table 2** versus 45.9% in **Table 4**) differ, but not significantly.

**Table 3.** Descriptive statistics including $PM_{10}$-traffic data

| | n | Mean [95% CI] | SD | Min. | Pctl (25) | Pctl (75) | Max. |
|---|---|---|---|---|---|---|---|
| Life expectancy [*years*] | 423 | 79.931 [79.811; 80.052] | 1.256 | 77.100 | 79.020 | 80.810 | 83.090 |
| Smokers [*share*] | 423 | 0.251 [0.248; 0.255] | 0.036 | 0.161 | 0.224 | 0.278 | 0.345 |
| Overweight [*share*] | 423 | 0.430 [0.425; 0.436] | 0.056 | 0.256 | 0.395 | 0.466 | 0.590 |
| $PM_{10}$-traffic [$\mu g/m^3$] | 423 | 26.623 [26.068; 27.177] | 5.802 | 14 | 22.2 | 29.5 | 49 |
| High school diploma [*percentage*] | 423 | 34.610 [33.712; 35.507] | 9.388 | 12 | 28.1 | 41.2 | 64 |
| Unemployment [*per 1,000 persons**] | 423 | 71.485 [68.660; 74.310] | 29.560 | 20.200 | 48.300 | 91.900 | 186.000 |
| Income [*1,000 Euros*] | 423 | 18,979.750 [18,695.14; 18,264.37] | 2,978.059 | 13,127 | 17,025.5 | 20,705.5 | 35,587 |
| Urban [*1=Urban*] | 423 | 0.740 [0.698; 0.782] | 0.439 | 0 | 0 | 1 | 1 |
| East [*1=East Germany*] | 423 | 0.248 [0.207; 0.290] | 0.432 | 0 | 0 | 0 | 1 |

**Table 4.** Descriptive statistics including $PM_{2.5}$-background data

| | n | Mean [95% CI] | SD | Min. | Pctl (25) | Pctl (75) | Max. |
|---|---|---|---|---|---|---|---|
| Life expectancy [*years*] | 143 | 80.498 [80.315; 80.680] | 1.104 | 77.860 | 79.810 | 81.065 | 83.600 |
| Smokers [*share*] | 143 | 0.236 [0.229; 0.243] | 0.041 | 0.124 | 0.214 | 0.262 | 0.415 |
| Overweight [*share*] | 143 | 0.459 [0.449; 0.469] | 0.061 | 0.304 | 0.422 | 0.499 | 0.609 |
| $PM_{2.5}$-background [$\mu g/m^3$] | 143 | 12.724 [12.330; 13.117] | 2.379 | 6 | 11.2 | 14 | 18 |
| High school diploma [*percentage*] | 143 | 36.109 [34.597; 37.621] | 9.148 | 15.200 | 29.450 | 41.850 | 59.100 |
| Unemployment [*per 1,000 persons**] | 143 | 58.358 [54.385; 62.331] | 24.036 | 17.900 | 40.300 | 74.750 | 118.400 |
| Income [*1,000 Euros*] | 143 | 20,590.710 [20,063.78; 21,117.63] | 3,187.519 | 15,191 | 18,500 | 22,031 | 35,587 |
| Urban [*1=Urban*] | 143 | 0.594 [0.513; 0.676] | 0.493 | 0 | 0 | 1 | 1 |
| East [*1=East Germany*] | 143 | 0.308 [0.231; 0.384] | 0.463 | 0 | 0 | 1 | 1 |

**Table 5.** Descriptive statistics including $PM_{2.5}$-traffic data

| | n | Mean [95% CI] | SD | Min. | Pctl (25) | Pctl (75) | Max. |
|---|---|---|---|---|---|---|---|
| Life expectancy [*years*] | 70 | 80.837 [80.619; 81.055] | 0.915 | 78.800 | 80.208 | 81.530 | 83.010 |
| Smokers [*share*] | 70 | 0.228 [0.220; 0.235] | 0.031 | 0.161 | 0.201 | 0.252 | 0.302 |
| Overweight [*share*] | 70 | 0.437 [0.425; 0.449] | 0.049 | 0.321 | 0.407 | 0.474 | 0.573 |
| $PM_{2.5}$-traffic [$\mu g/m^3$] | 70 | 14.471 [13.903; 15.040] | 2.385 | 10 | 13 | 16 | 23 |
| High school diploma [*percentage*] | 70 | 38.131 [36.068; 40.195] | 8.654 | 13.700 | 32.400 | 43.600 | 57.900 |
| Unemployment [*per 1,000 persons**] | 70 | 51.729 [46.956; 56.502] | 20.019 | 20.200 | 36.025 | 65.525 | 120.200 |
| Income [*1,000 Euros*] | 70 | 21,360.470 [20,537.77; 22,183.17] | 3,450.329 | 15,849 | 19,056.5 | 22,640.5 | 35,587 |
| Urban [*1=Urban*] | 70 | 0.700 [0.590; 0.810] | 0.462 | 0 | 0 | 1 | 1 |
| East [*1=East Germany*] | 70 | 0.243 [0.140; 0.346] | 0.432 | 0 | 0 | 0 | 1 |

The smallest database is for $PM_{2.5}$-traffic pollution. Comparing **Table 5** with its equivalent **Table 3** ($PM_{2.5}$-background pollution) reveals especially a significantly lower share of smokers.

## Estimation Approach

Data matching is the applied here, since this method seems appropriate for the data structure described above. Moreover, as the foregoing brief literature review revealed, the differences in life expectancy and PM air pollution at the district level are considerable in Germany. In accordance with [35], matching methods are preferable to other methods in observational studies like this one, because they allow for a more precise determination of groups with similar covariates, when estimating the differences between the "treatment effects" (smoking, overweight, and PM air pollution) on life expectancy. In this way, (intentional or unintentional) biases to obtain preferred results can also be prevented [35, 36].

Another issue with OLS estimations (and many others, for instance, logit, probit, and count data models) is that they are parametric methods [37]. The estimation equation must therefore be specified and can be changed very easily and very quickly. Thus, the estimations are model-dependent [37]. In this respect, matching methods seem to be probably a better choice, as the treatment group is more rigorously determined, without prespecifying a parametric model.

The employed matching method (see [38] for an overview) proceeds, as follows. First, each county is associated with another county, according to the criterion that the Euclidean (or the Manhattan) distance concerning the control variables is minimized. The Euclidean distance, $d_E(\boldsymbol{x}, \boldsymbol{y})$, with $\boldsymbol{x}$ and $\boldsymbol{y}$ as vectors of variables, is defined by [39]:

$$d_E(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}, \quad (1)$$

and the Manhattan distance, $d_M(\boldsymbol{x}, \boldsymbol{y})$, by [40]:

$$d_M(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n}|x_i - y_i|. \quad (2)$$

The reason for applying different distance measures for the county matching is that outliers obtain a heavy weight in Euclidean distance, whereas they are treated equally in Manhattan distance. Both distance measures are applied to control whether these differences in weighting may cause different results.

Since the control variables are measured in different units, they are normalized. The reason for matching counties in this way is that they should be as similar as possible with respect to the controls. This matching method is applied each year with observations, i.e., for the years 1999, 2003, 2005, 2009, 2013, and 2017. For example, to analyze the relationship between life expectancy and smoking, the respective control

**Table 6.** Correlation of shares of smokers and overweight people with life expectancy for matched district-level data, without controls for particulate matter $PM_{2.5}/PM_{10}$ concentration

| Correlation with life expectancy | # of matches | # of county-year-values | # of different included counties | EC [95% CI] |
|---|---|---|---|---|
| Share of smokers | 2,197 | 2,197 | 394 | -0.1729*** [-0.213; -0.132] |
| Share of overweight people | 2,197 | 2,197 | 394 | -0.2676*** [-0.306; -0.228] |

Note. *p<0.1; **p<0.05; ***p<0.01; EC: Estimation coefficient; & Source: Own calculations

**Table 7.** Correlation for particulate matter $PM_{2.5}/PM_{10}$ concentration with life expectancy for matched district-level data, without controls for shares of smokers and overweight people

| Correlation with life expectancy | # of matches | # of county-year-values | # of different included counties | EC [95% CI] |
|---|---|---|---|---|
| Background $PM_{2.5}$ concentration | | | | |
| Euclidean distance | 31 | 39 | 35 | -0.0455 [-0.393; 0.314] |
| Manhattan distance | 27 | 36 | 30 | 0.1344 [-0.259; 0.490] |
| Traffic $PM_{2.5}$ concentration | | | | |
| Euclidean distance | 10 | 13 | 13 | 0.0937 [-0.570; 0.683] |
| Manhattan distance | 10 | 15 | 15 | 0.1368 [-0.539; 0.706] |
| Background $PM_{10}$ concentration | | | | |
| Euclidean distance | 510 | 584 | 197 | 0.0942** [0.007; 0.180] |
| Manhattan distance | 497 | 571 | 194 | 0.1059** [0.019; 0.192] |
| Traffic $PM_{10}$ concentration | | | | |
| Euclidean distance | 223 | 252 | 97 | 0.0141 [-0.117; 0.145] |
| Manhattan distance | 223 | 250 | 98 | 0.0952 [0.034; 0.229] |

Note. *p<0.1; **p<0.05; ***p<0.01; EC: Estimation coefficient; & Source: Own calculations

variables are the share of overweight persons, the concentration of fine dust, the rate of unemployment, disposable income per capita, the rate of high school graduates, as well as dummy variables for urban counties and Eastern Germany. Note that only such counties are considered for which the set of control variables is complete.

In a second step, for each county match, the differences in life expectancies on the one hand, and the difference in the rate of smokers (rate of overweight persons, $PM_{2.5}$ or $PM_{10}$ pollution, respectively) in comparable units on the other are determined. Consider the following fictious example. In 2009, county F was matched with county L. Life expectancy in county F was 82 years and in county L 80 years, which yields a difference of 2. The share of smokers in county F was 10 units and 12 units in county L, yielding a smoker difference of -2.

In a third step, the differences from step two for all observation years are combined in one data frame, where the first column contains the differences in life expectancies and the second column the differences in smoker shares. With this data, the correlation coefficient between life expectancy and smoker shares is calculated, as well as the statistical significances. In the following, the level for statistical significance is set at an error level of 5%, indicated by ** asterisks. Note that the error levels are also given for 10% by * asterisk and 1% by *** asterisks.

# RESULTS OF THE DATA MATCHING ANALYSIS

## Estimations Without Respective Controls for PM Air Pollution and Shares of Smokers and Overweight Persons

Below, the correlation between the explanatory variables (share of smokers and overweight people, as well as $PM_{2.5}$ and $PM_{10}$ pollution concentration) and life expectancy at birth are presented. Thereby, the estimation was run according to the method described before using the statistical software R.

The results shown in **Table 6** are estimations without PM pollution concentration, in order to provide a benchmark for the following estimations that include PM pollution. Note that all estimations rely on county-level matched data, for which the above control variables are used for the matching. Furthermore, the results are the same for Euclidean and Manhattan distance. Controlling for overweight, the share of people with a high school diploma, the number of unemployed per 1,000 people of working age, disposable income, rural or urban location of living and a dummy variable for East Germany, an increase of the share of smokers evidently decreases life expectancy at birth. In addition, running the estimation for the share of overweight people shows that an increase in the share of overweight people decreases also life expectancy at birth. Both results are statistically significant at the 1% error level. The 95%-confidence intervals show that also these intervals differ only slightly.

In contrast to the results in **Table 6**, **Table 7** shows disappointing results for the matched correlations between fine particulate matter $PM_{2.5}$ and $PM_{10}$ with life expectancy. The number of $PM_{2.5}$ (background and traffic) recording stations is seemingly too small for a statistical analysis with a matching method. Only 10 to 31 matches were possible. Moreover, all estimated correlations are highly statistically insignificant.

In the lower part of **Table 7**, the estimations for $PM_{10}$ air pollution are shown. A higher density of recording stations allows for many more matches. However, the number of matches is still considerably lower than for the shares of smokers and overweight people. But as in the case of $PM_{2.5}$, all but one of the estimated correlations are statistically insignificant.

**Table 8.** Correlation of share of smokers, share of overweight people and level of background $PM_{2.5}/PM_{10}$ with life expectancy with matched district-level data (Euclidean distance)

| Correlation with life expectancy | # of matches | # of county-year-values | # of different included counties | EC [95% CI] |
|---|---|---|---|---|
| Results for background $PM_{2.5}$ | | | | |
| Share of smokers | 143 | 143 | 81 | -0.2652*** [-0.411; -0.106] |
| Share of overweight people | 143 | 143 | 81 | -0.5518*** [-0.656; -0.426] |
| Level of background $PM_{2.5}$ | 32 | 44 | 36 | -0.1283 [-0.456; 0.231] |
| Results for background $PM_{10}$ | | | | |
| Share of smokers | 826 | 826 | 217 | -0.2710*** [-0.333; -0.207] |
| Share of overweight people | 826 | 826 | 217 | -0.3615*** [-0.419; -0.301] |
| Level of background $PM_{10}$ | 466 | 540 | 199 | -0.0546 [-0.145; 0.036] |

Note. *p<0.1; **p<0.05; ***p<0.01; EC: Estimation coefficient; & Source: Own calculations

**Table 9.** Correlation of share of smokers, share of overweight people, and level of background PM2.5/PM10 with life expectancy with matched district-level data (Manhattan distance)

| Correlation with life expectancy | # of matches | # of county-year-values | # of different included counties | EC [95% CI] |
|---|---|---|---|---|
| Results for background $PM_{2.5}$ | | | | |
| Share of smokers | 143 | 143 | 81 | -0.267*** [-0.413; -0.107] |
| Share of overweight people | 143 | 143 | 81 | -0.5892*** [-0.687; -0.471] |
| Level of background $PM_{2.5}$ | 31 | 40 | 34 | 0.071 [-0.291; 0.415] |
| Results for background $PM_{10}$ | | | | |
| Share of smokers | 826 | 826 | 217 | -0.2624*** [-0.325; -0.198] |
| Share of overweight people | 826 | 826 | 217 | -0.3263*** [-0.325; -0.198] |
| Level of background $PM_{10}$ | 449 | 518 | 196 | 0.008 [-0.100; 0.085] |

Note. *p<0.1; **p<0.05; ***p<0.01; EC: Estimation coefficient; & Source: Own calculations

**Table 10.** Correlation of share of smokers, share of overweight people, and level of traffic $PM_{2.5}/PM_{10}$ with life expectancy with matched district-level data (Euclidean distance)

| Correlation with life expectancy | # of matches | # of county-year-values | # of different included counties | EC [95% CI] |
|---|---|---|---|---|
| Results for traffic $PM_{2.5}$ | | | | |
| Share of smokers | 70 | 70 | 45 | -0.1920 [-0.409; 0.045] |
| Share of overweight people | 70 | 70 | 45 | -0.6339*** [-0.756; -0.469] |
| Level of traffic $PM_{2.5}$ | 15 | 17 | 17 | -0.1805 [-0.634; 0.366] |
| Results for traffic $PM_{10}$ | | | | |
| Share of smokers | 423 | 423 | 127 | -0.3220*** [-0.405; -0.234] |
| Share of overweight people | 423 | 423 | 127 | -0.3759*** [-0.455; -0.291] |
| Level of traffic $PM_{10}$ | 209 | 238 | 94 | 0.0691 [-0.067; 0.203] |

Note. *p<0.1; **p<0.05; ***p<0.01; EC: Estimation coefficient; & Source: Own calculations

The statistically significant correlation–background $PM_{10}$ concentration with the Euclidean distance measure–has an unexpected and implausible positive sign.

The results concerning the directed correlations between the shares of smokers/overweight persons and life expectancy is in accordance with the results for Germany [9], as well as other countries [27, 28]. This can be taken as an indication that this relationship does not depend on the applied empirical estimation method, i.e., it seems to be method-invariant.

### Estimations Including Respective Controls for PM Air Pollution and Shares of Smokers and Overweight People

In **Table 8**, the results for estimations including PM background concentration are shown, based on the Euclidean distance metric. Whereas the results for share of smokers and share of overweight people (including PM-concentration as control variables) are structurally equal to those in **Table 6**, the correlation between life expectancy and $PM_{2.5}$, as well as $PM_{10}$, concentration is positive, but statistically insignificant. The conclusion is, therefore, that in this matching model PM background pollution has no effect on life expectancy. By contrast, the strong negative effects of smoking and overweight on life expectancy is confirmed–statistically highly significantly. Moreover, by controlling for PM pollution, the negative effects of smoking and overweight are considerably stronger than in the estimation without these controls in **Table 6**.

The estimation results documented in **Table 9**, which are based on the Manhattan distance metric, do not confirm the former results in all respects. With the smallest dataset (n=70) of $PM_{2.5}$ pollution, only the share of overweight people has a strong, statistically significant, negative effect on life expectancy. Neither the share of smoking people, nor $PM_{2.5}$ background pollution, have such an impact. However, this is different with $PM_{10}$ background pollution, for which the database is much larger. Smoking as well as overweight have statistically significant negative effects on life expectancy, which are again greater than these effects without $PM_{10}$ control in **Table 6**. $PM_{10}$ background is not statistically significant.

Very similar results for the same estimations with the Euclidean distance metric for traffic-related PM pollution are found and presented in **Table 10**. No new insights are gained there.

**Table 11.** Correlation of share of smokers, share of overweight people, and level of traffic $PM_{2.5}$/$PM_{10}$ with life expectancy with matched district-level data (Manhattan distance)

| Correlation with life expectancy | # of matches | # of county-year-values | # of different included counties | EC [95% CI] |
|---|---|---|---|---|
| Results for traffic $PM_{2.5}$ | | | | |
| Share of smokers | 70 | 70 | 45 | -0.1478 [-0.370; 0.090] |
| Share of overweight people | 70 | 70 | 45 | -0.5999*** [-0.732; -0.425] |
| Level of traffic $PM_{2.5}$ | 15 | 19 | 18 | -0.2785 [-0.692; 0.273] |
| Results for traffic $PM_{10}$ | | | | |
| Share of smokers | 423 | 423 | 127 | -0.3220*** [-0.405; -0.234] |
| Share of overweight people | 423 | 423 | 127 | -0.3759*** [-0.455; -0.291] |
| Level of traffic $PM_{10}$ | 209 | 238 | 94 | 0.0691 [-0.067; 0.203] |

Note. *p<0.1; **p<0.05; ***p<0.01; EC: Estimation coefficient; & Source: Own calculations

Applying the Manhattan distance metric for traffic PM pollution yields the results in **Table 11**. With $PM_{10}$, smoking and obesity have a negative statistically significant effect on life expectance, whereas PM traffic pollution does not have such an effect. With the smaller $PM_{2.5}$ dataset, only obesity shows a statistically significant impact. The effects are very similar to those with the Euclidean distance measure demonstrated in **Table 10**.

The results presented in this subsection are remarkable with regard to PM air pollution. The medical mechanism for PM's damage to the human body [20-24], are very well documented, as are the negative long-term effects of PM air pollution on life expectancy [41, 42]. The latter was also shown for Germany (see, for instance, [5, 15]. For this reason, it is surprising that it was not possible to find a conclusive negative relationship between long-term PM air pollution and life expectancy in this study so far. By contrast, the negative relationships between smoking/overweight on life expectancy were confirmed in this study. There are two possible explanations for this result. The first explanation is the lack of measuring stations in Germany. However, this study and the studies of Prinz and Richter used the same database. The second explanation is that the relationship between PM air pollution and life expectancy is not method-invariant, in contrast to the results for smoking and overweight. A plausible reason for these contrasting results is the considerable smaller number of matches and included counties for the PM estimations. Nevertheless, this smaller number is a consequence of the applied matching method.

In the next section, OLS estimations are applied as robustness checks for the results of the matching analysis. Moreover, it is studied whether PM air pollution is a moderator variable concerning the relationship of smoking/overweight and life expectancy.

## ROBUSTNESS CHECKS AND EXTENSIONS WITH OLS ESTIMATIONS

In this section, we determine whether the results with the above matching method are robust when OLS estimations are applied. Although in several respects, matching seems to be somewhat better suited for the data analysis, it is nevertheless useful to analyze the data with a different method. In particular, we test with the OLS estimations below, whether the lack of statistically significant evidence of a negative effect of PM air pollution on life expectancy depends on the estimation method, i.e., that this result is not robust with respect to the estimation approach. Furthermore, OLS is used to estimate whether there is a moderator effect of PM pollution on the impact of smoking and overweight on life expectancy.

The OLS estimation equation reads:

$$E = \beta_0 + \beta_1 \cdot SMO(OVW) + \beta_2 \cdot PM_i + \beta_3 \cdot [SMO(OVW) \cdot PM_i] + \sum_{j=4}^{n} \beta_j CON_j + \varepsilon, \tag{3}$$

$LE$: Life expectancy,

$SMO$: Share of smokers,

$OVW$: Share of overweight people,

$PM_i$: Particulate matter pollution, $i = \{10; 2.5\}$,

$CON_j$: Control variable $j$,

$\beta_j$: Estimation coefficient, $\beta_j = 0, 1, \ldots, n$, and

$n$: Number of explanatory variables.

Note that $\beta_3 \cdot [SMO(OVW) \cdot PM_i]$ measures the moderating effect of $PM_i$ on the impact of $SMO$ and $OVW$, respectively, on life expectancy, $LE$. In addition, the first derivative of equation (3) with respect to SMO and OVW, respectively, gives:

$$\frac{\partial LE}{\partial SMO(OVW)} = \beta_1 + \beta_3 \cdot PM_i. \tag{4}$$

For $PM_i = 0$, this is the direct effect of SMO and OVW, respectively, on LE. Expressed differently, $\beta_1$ is the slope of the function LE (SMO, OVW; $PM_i = 0$). Differentiating equation (4) with respect to $PM_i$ yields:

$$\frac{\partial^2 LE}{\partial SMO(OVW)\partial PM_i} = \beta_3. \tag{5}$$

The estimation coefficient $\beta_3$ determines how the slope of LE (SMO, OVW; $PM_i$) is changed by changes in $PM_i$. In other words, $\beta_3$ measures the moderating effect of particulate matter pollution on the impact of SMO and OVW, respectively, on LE. If $\beta_3 \neq 0$ and statistically significant, particulate matter pollution can be considered a moderator of the effect of smoking and overweight, respectively, on life expectancy. For $\beta_3 < (>) 0$, the slope of the function LE (SMO, OVW; $PM_i$) decreases (increases).

**Table 12** shows the OLS estimation results with background PM air pollution. The columns 1 and 2 in **Table 12** are estimations with smoking and overweight, respectively, as the impact-variables of interest with $PM_{10}$ pollution. Columns 3 and 4 contain the estimation results for $PM_{2.5}$ pollution. The $PM_{10}$ estimations have better statistical properties with higher values of the F-statistic (it tests whether the estimation coefficients of all explanatory variables are simultaneously equal to zero) and higher values for explained variances, i.e.,

**Table 12.** OLS estimations for smoking and overweight on life expectancy, with background PM pollution

| Background PM | Dependent variable: Life expectancy | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Smoker (share) | -14.205*** (2.903) | | -4.874 (8.222) | |
| Overweight (share) | | -15.805*** (2.331) | | 0.741 (5.577) |
| $PM_{10}$ | -0.092*** (0.035) | -0.238*** (0.047) | | |
| $PM_{2.5}$ | | | -0.029 (0.159) | 0.351* (0.198) |
| High school diploma | 0.013*** (0.004) | 0.001 (0.004) | 0.015** (0.007) | -0.004 (0.007) |
| Unemployment per 1,000 | -0.020*** (0.002) | -0.021*** (0.002) | -0.029*** (0.004) | -0.023*** (0.004) |
| Income (1,000 Euro) | 0.0001*** (0.00002) | 0.0001*** (0.00002) | 0.0001* (0.00004) | 0.0001** (0.00003) |
| Urban | 0.445*** (0.088) | 0.277*** (0.083) | 0.623*** (0.164) | 0.291* (0.155) |
| East Germany | 0.735*** (0.118) | 1.013*** (0.116) | 0.612*** (0.164) | 0.810*** (0.167) |
| Smoker · $PM_{10}$ | 0.343*** (0.128) | | | |
| Overweight · $PM_{10}$ | | 0.521*** (0.108) | | |
| Trend | 55.475*** (5.106) | 64.052*** (5.298) | | |
| $Trend^2$ | -42.087*** (2.180) | -36.535*** (2.159) | | |
| Smoker · $PM_{2.5}$ | | | 0.135 (0.611) | |
| Overweight · $PM_{2.5}$ | | | | -0.719* (0.437) |
| Constant | 81.053*** (0.993) | 85.077*** (1.180) | 80.639*** (2.645) | 79.453*** (2.754) |
| Observations | 826 | 826 | 143 | 143 |
| Adjusted $R^2$ | 0.768 | 0.777 | 0.643 | 0.728 |
| F statistic | 274.503*** | 288.232*** | 32.931*** | 48.614*** |
| AIC | 1,534.873 | 1,503.618 | 297.682 | 258.451 |

Note. *p<0.1; **p<0.05; ***p<0.01; & Source: Own calculations

higher values of adjusted $R^2$. For the $PM_{10}$ and $PM_{2.5}$ estimations, the results for the overweight variable are statistically more reliable than those for the smoker variable, as can be seen by higher adjusted $R^2$ values, as well as and lower Akaike information criterion (AIC) values, for the overweight estimations. The AIC is especially used for model selection, whereby a lower AIC value means a better model. According to the AIC, an estimation model is better than another version of the model if it explains more of the variance with less independent variables (see for technical details in [43]).

Although both smoking and overweight are negatively and statistically significantly correlated with life expectancy, the effect is somewhat stronger for overweight. The covariates are all also statistically significant with the expected signs: a positive correlation with life expectancy for people with higher disposable incomes and a negative correlation with the unemployment level. Living in an urban environment and in East Germany has a positive correlation with life expectancy. Finally, life expectancy has a statistically significant positive trend that is decreasing over time, due to the negative sign of the squared trend variable.

In contrast to the matching results above, the OLS estimations show that $PM_{10}$ pollution has a direct statistically significantly negative effect on life expectancy, besides the negative impact of smoking and overweight. This effect is stronger in the overweight estimation than in the smoker estimation. Interestingly, the share of people with high school diplomas loses statistical significance in the overweight estimation, but not in the smoker estimation.

In addition, $PM_{10}$ pollution has a moderating effect on the impact of smoking and overweight on life expectancy. The interaction effects–smoker · $PM_{10}$ and overweight · $PM_{10}$–are statistically significant at the 1% error level and they have a positive sign. Although the direct $PM_{10}$ pollution effects on life expectancy are statistically significantly negatively correlated with life expectancy, the combined effect with smoking and overweight (also negatively correlated with life expectancy) is positive. This can be interpreted as meaning that the indirect moderation effect attenuates the impact of the negative direct effects on life expectancy.

The effects of $PM_{2.5}$ in columns 3 and 4 of **Table 12** are either not statistically significant or only weakly significant, but with an unexpected sign. These results come with a high degree of uncertainty since the number of observations, 143, is much smaller than the observations with $PM_{10}$ data, 826. Therefore, these estimation results are only shown for the sake of completeness. Moreover, the OLS estimations with traffic PM pollution are delegated to the **Appendix A** as **Table A1**, **Figure A1**, and **Figure A2**.

In **Figure 2** and **Figure 3**, the moderator effects of background $PM_{10}$ and $PM_{2.5}$ air pollution is shown graphically. **Figure 2** presents the effect for the smoker variable. In the upper part of **Figure 2**, the middle of the three lines shows the mean effect of $PM_{10}$ on the relationship between the share of smokers and life expectancy, whereas (from left to right) the upper (lower) line contains the effect of a one standard deviation lower (higher) $PM_{10}$ concentration. Mean $PM_{10}$ concentration has a lowering effect on life expectancy, as shown above, and the effect is the stronger, the higher the share of smokers is. However, the moderator effect of plus (minus) one standard deviation of $PM_{10}$ concentration reduces (increases) the slope of the regression line with higher (lower) shares of smokers. Moreover, the lower part of **Figure 2** shows the moderating effect of $PM_{2.5}$. All three regression lines are very close to each other, which represents the fact that there is no statistically significant moderating effect of $PM_{2.5}$ on the relationship smoking and live expectancy.

**Figure 3** shows the same effect of PM as a moderator variable, but on the relationship between overweight and life expectancy. Again, the respective middle line in the upper and lower part of **Figure 3** presents the mean effect of $PM_{10}$ and

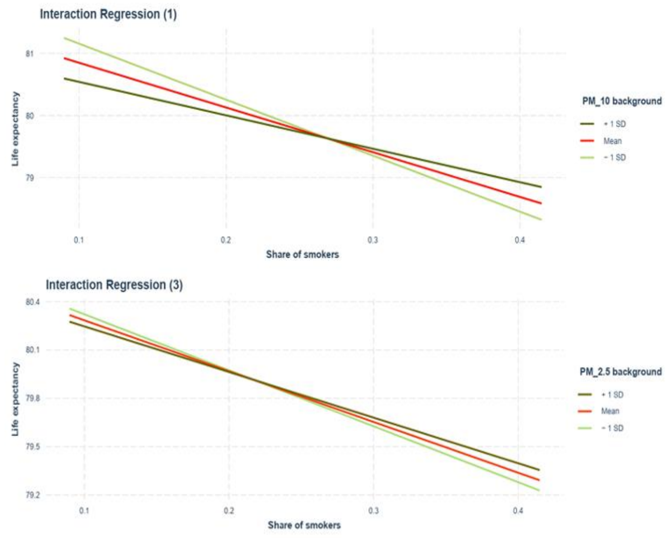


**Figure 2.** Moderator effect of background PM on the relationship between the share of smokers and life expectancy (Source: Own depiction with software R)

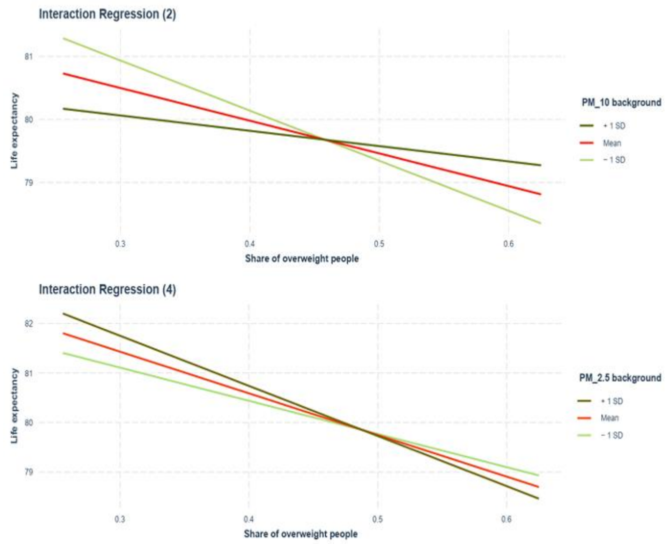


**Figure 3.** Moderator effect of background PM on the relationship between the share of overweight people and life expectancy (Source: Own depiction with software R)

$PM_{2.5}$ air pollution, respectively, on said relationship. In comparison with the moderating effect on the relationship between smoking and life expectancy, this effect is stronger concerning overweight and life expectancy. A one standard deviation higher (lower) $PM_{10}$ concentration decreases (increases) the slope of the respective regression line. However, this is vice versa in the lower part of **Figure 3**, i.e., for $PM_{2.5}$. According to the assumed error level for statistical significance of 5%, the moderator effect is statistically insignificant. The question whether this effect exists requires a considerably larger data base of $PM_{2.5}$ measurements.

## DISCUSSION

The empirical results in this paper confirm those concerning the negative effects of smoking and overweight on life expectancy, from studies mentioned in the introduction section. By contrast, the results on the effects of particulate matter air pollution differ greatly from other studies (see, for instance, [5, 15], as well as the literature quoted therein). In the empirical estimation with the matching model, neither $PM_{10}$ and $PM_{2.5}$ nor background PM and traffic-related PM pollution have a statistically significant negative effect on life expectancy. The reason for this difference cannot be found in the data base for PM pollution, as the data used in this paper are the same as those applied by [5, 15]. Therefore, the conclusion can be drawn that the estimation method seems to be decisive. As stated before, there is reason to assume that the applied matching method is better suited for the empirical than multiple regressions. However, the OLS estimation results show that the applied estimation method makes a difference. For the time being, a definitive answer to the question of which method gives the ‘correct’ results, does not seem possible.

This paper is an ecological study and not a clinical trial. Nevertheless, it is difficult (if not impossible) to study long-run effects with PM pollution in clinical trials with individual data. The approach in this paper is thus

(a) to use relevant control variables and

(b) to match data on the county level (and to check the robustness of the estimates with OLS estimations).

The county is the smallest administrative unit for which data are available in Germany.

An ongoing issue for German studies on PM pollution is the density of PM recording stations over Germany. A complete and even covering of entire Germany is, therefore, not possible. This issue can only be overcome with satellite data. However, it is not clear whether or to what extent the data gaps are relevant for the empirical results in this paper. Moreover, no individual data concerning smoking and overweight is available. As a novelty, in this paper, data of the German micro census are used for the estimations. Although this is not a perfect substitute for individual data, it is statistically a viable procedure when individual data is not available.

The most interesting question is, however, how the considerably stronger negative effects of smoking and overweight on life expectancy in estimations with PM pollution controls may be interpreted. An initial approach could be that the effect is a consequence of the smaller number of observations and matchings. By contrast, a second interpretation is that PM pollution acts as a moderator in these estimations. As a moderator variable, PM pollution can change the impact of smoking and overweight on life expectancy, even if the variable itself is not statistically significant.

As shown in **Figure 1**, environmental variables, and even more importantly, the local environmental situation, is in general considered relevant for the relationship between health behavior or lifestyles and life expectancy. In the specific case of smoking and overweight, air pollution could be an even more important of the local quality of the environment. This reasoning is in accordance with the OLS estimation results, as the interaction effects for smoking (and overweight) and $PM_{10}$ are statistically significant and larger than zero.

If the second interpretation is the correct one, local PM air pollution moderates the effect of smoking and overweight in such a way that for a given level of PM concentration, the effects of smoking and overweight become more obvious. Nevertheless, the first interpretation of the results, that the relatively small number of pollution-recording stations is the true reason for the results of the matching model, cannot be excluded without better data.

This leads immediately to the limitations of the study. Firstly, German data on $PM_{2.5}$ are only available for quite short period of time. Therefore, statistical significance might be underestimated because of the small data base. Secondly, the distribution of PM measuring stations over Germany is unequal. Therefore, it is not clear whether or to what an extent the empirical results of the paper are biased by the measuring station distribution. Thirdly, there is no data available about the mobility of people over their lifetimes. A person that lives at the time of the Micro Census at a certain place could have lived at other places before. The Micro Census data do not contain information on that. Fourthly, additional information on socio-economic characteristics of the Micro Census interviewee are missing, e.g., health status at the time of the interview and age.

In our view, the first focus of future research should be on employing satellite data on PM air pollution [44, 45] that avoids any distribution bias of measuring stations. This would give a much better picture of PM air pollution. The second focus should be on persons who lived their entire life in one place (or in a short distance to the place where they were raised). This would exclude migration effects. Moreover, on should focus on the health status of people at the time of the Micro Census interview.

## CONCLUSION

In this paper, the effects of smoking, overweight and PM air pollution on life expectancy in German counties are studied. German Micro Census data are used to measure the distribution of smoking and overweight in Germany. $PM_{10}$ and $PM_{2.5}$ fine particulate matter air pollution data at the county level from all measuring stations throughout Germany are used in this paper, in addition to the Micro Census data. A second novelty with respect to existing studies is the applied estimation method. In this paper, a matching approach is chosen. The main reason for this choice is that multiple regression analysis–which is applied in other studies–is a parametric method requiring model prespecification. This is different with a matching approach. The latter is a non-parametric method for analyzing data by controlling more rigorously for covariates. Furthermore, two different distance metrics are used, the Euclidean distance and the Manhattan distance. Nonetheless, we employ OLS estimations as a robustness check in this paper.

The results of smoking and overweight on life expectancy from other studies are confirmed in this study. Both reduce life expectancy considerably, with overweight being more serious than smoking according to this analysis. In contrast to other studies on long-term PM air pollution effects in Germany on life expectancy, in this study, no relevant negative and statistically significant effect was found with the matching method. However, when using PM pollution as a covariate in the estimations of smoking and overweight effects on life expectancy, the negative effects of smoking and overweight become more obvious and larger. One interpretation of this result is that PM air pollution moderates the effect of smoking and overweight on life expectancy. This was confirmed by OLS-regressions with the same data. In contrast to the matching estimations, OLS estimations provided evidence of statistically significant and negative direct effects of background $PM_{10}$ pollution on life expectancy. Moreover, the interaction effects of $PM_{10}$ pollution and smoking, as well as overweight, were statistically significant, indicating an additional indirect moderating effect of $PM_{10}$ pollution that attenuates the direct negative effects on life expectancy.

**Author contributions:** All co-authors have involved in all stages of this study while preparing the final version. They all agree with the results and conclusions.

**Funding:** No funding source is reported for this study.

**Declaration of interest:** No conflict of interest is declared by the authors.

**Ethical statement:** Authors stated that the study used routinely-collected and anonymized official data of the Federal Statistical Office and the Statistical Offices of the Länder, as well as the Federal Environmental Agency of Germany.

**Data sharing statement:** Data supporting the findings and conclusions are available upon request from corresponding author.

## REFERENCES

1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1223-49. https://doi.org/10.1016/S0140-6736(20)30752-2 PMid:33069327
2. Rizzuto D, Fratiglioni L. Lifestyle factors related to mortality and survival: A mini-review. Gerontology. 2014;60(4):327-35. https://doi.org/10.1159/000356771 PMid:24557026
3. Keeney RL. Personal decisions are the leading cause of death. Oper Res. 2008;56(6):1335-47. https://doi.org/10.1287/opre.1080.0588
4. Finkelstein A, Gentzkow M, Williams H. Place-based drivers of mortality: Evidence from migration. Am Econ Rev. 2021;111(8):2697-735. https://doi.org/10.1257/aer.20190825 PMid:34887592 PMCid:PMC8653912
5. Prinz A, Richter DJ. Feinstaubbelastung und Lebenserwartung in Deutschland [Particulate matter pollution and life expectancy in Germany]. AStA Wirtsch Sozialstat Arch. 2021;15:237-72. https://doi.org/10.1007/s11943-021-00292-1
6. Barabási A-L. Network medicine—from obesity to the "diseasome". N Engl J Med. 2007;357(4):404-7. https://doi.org/10.1056/NEJMe078114 PMid:17652657

7. Goh K-I, Cusick ME, Valle D, Barabási A-L. The human disease network. PNAS. 2007;104(21):8685-90. https://doi.org/10.1073/pnas.0701361104 PMid:17502601 PMCid:PMC1885563
8. Loscalzo J, Kohane I, Barabási A-L. Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol. 2007;3:124. https://doi.org/10.1038/msb4100163 PMid: 17625512 PMCid:PMC1948102
9. Nussfelder WJ, Valverde JR. Impact of individual risk factors on German life expectancy. Erasmus MC, Rotterdam. 2018. Available at: https://www.gdv.de/resource/blob/35298/1b80120a041c3ef97315726ac9ebe226/download-studie-lebenszeitkiller-rotterdam-data.pdf (Accessed 20 January 2022).
10. Lampert T, Hoebel J, Kroll LE. Social differences in mortality and life expectancy in Germany. Current situation and trends. J Health Monit. 2019;4(1):3-14. https://doi.org/10.25646/5872 PMid:35586309 PMCid: PMC8848850
11. Lampert T, Kroll LE, von der Lippe E, Müters S, Stolzenberg H. Sozioökonomischer Status und Gesundheit [Socio-economic status and health]. Bundesgesundheitsblatt. 2013;56:814-21. https://doi.org/10.1007/s00103-013-1695-4 PMid:23703503
12. Chetty R, Stepner M, Abraham S, et al. The association between income and life expectancy in the United States, 2001-2014. JAMA. 2016;315(16):1750-66. https://doi.org/10.1001/jama.2016.4226 PMid:27063997 PMCid: PMC4866586
13. Wengler A, Rommel A, Plaß D, et al. Years of life lost to death. Dtsch Arztebl Int. 2021;118(9):137-44. https://doi.org/10.3238/arztebl.m2021.0148 PMid: 33958031 PMCid:PMC8212398
14. Rau R, Schmertmann CP. District-level life expectancy in Germany. Dtsch Arztebl Int. 2020;117(29-30):493-9. https://doi.org/10.3238/arztebl.2020.0493 PMid:33087229 PMCid:PMC7588608
15. Prinz AL, Richter DJ. Long-term exposure to fine particulate matter and air pollution: An ecological study of its effect on COVID-19 cases and fatality. Environ Res. 2022;204(Pt A):111948. https://doi.org/10.1016/j.envres.2021.111948 PMid:34464613 PMCid:PMC8400616
16. von Bertalanffy L. The theory of open system in physics and biology. Sci. 1950;111(2872):23-9. https://doi.org/10.1126/science.111.2872.23 PMid:15398815
17. The Academy of Medical Sciences. Identifying the environmental causes of disease: How should we decide what to believe and when to take action? Academy of Medical Sciences. 2007. Available at: https://acmedsci.ac.uk/file-download/34654-119615475058.pdf (Accessed 20 January 2022).
18. Hill AB. The environment and disease: Association or causation? Proc R Soc Med. 1965;58(5):295-300. https://doi.org/10.1177/003591576505800503 PMid: 14283879 PMCid:PMC1898525
19. Jaeger H. Multifunctionality: A fundamental property of behavior mechanisms based on dynamical systems. In: Pfeifer R, Blumberg B, Meyer J-A, Wilson SW, editors. From animals to animats 5: Proceedings of the 5th International Conference on Simulation of Adaptive Behavior. Cambridge, MA: MIT Press; 1998. p. 286-90.
20. Pope III, CA. Epidemiology of fine particulate air pollution and human health: Biological mechanism and who's at risk? Environ Health Perspect. 2000;108(Suppl 4):713-23. https://doi.org/10.1289/ehp.108-1637679 PMid:10931790 PMCid:PMC1637679
21. Anderson JO, Thundiyil JG, Stolbach A. Clearing the air: A review of the effects of particulate matter air pollution on human health. J Med Toxicol. 2012;8(2):166-75. https://doi.org/10.1007/s13181-011-0203-1 PMid: 22194192 PMCid:PMC3550231
22. Hamanaka RB, Mutlu GM. Particular matter air pollution: Effects on the cardiovascular system. Front Endocrinol (Lausanne). 2018;9:680. https://doi.org/10.3389/fendo.2018.00680 PMid:30505291 PMCid:PMC6250783
23. Losacco C, Perillo A. Particulate matter air pollution and respiratory impact on humans and animals. Environ Sci Pollut Res. 2018;25(34):33901-10. https://doi.org/10.1007/s11356-018-3344-9 PMid:30284710
24. Lelieveld S, Wilson J, Dovrou E, et al. Hydroxyl radical production by air pollutants in epithelial lining fluid governed by interconversion and scavenging of reactive oxygen species. Environ Sci Technol. 2021;55(20):14069-79. https://doi.org/10.1021/acs.est.1c03875 PMid: 34609853 PMCid:PMC8529872
25. Braun M, Fromm E-L, Gerber A, Klingelhöfer D, Müller R, Groneberg DA. Particulate matter emission of four types of one cigarette brand with and without additives: A laser spectrometric particulate matter analysis of secondhand smoke. BMJ Open. 2019; 9(1):e024400. https://doi.org/10.1136/bmjopen-2018-024400 PMid:30782733 PMCid: PMC6340634
26. Bhaskaran K, Dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: A population-based cohort study 3.6 million adults in the UK. Lancet Diabetes Endocrinol. 2018;6(12):944-53. https://doi.org/10.1016/S2213-8587(18)30288-2 PMid:30389323
27. Darden M, Gilleskie DB, Strumpf K. Smoking and mortality: New evidence from a long panel. Int Econ Rev (Philadelphia). 2018;59(3):1571-619. https://doi.org/10.1111/iere.12314 PMid:31274880 PMCid:PMC6608712
28. Janssen F. The role of smoking in country differences in life expectancy across Europe, 1985-2014. Nicotine Tob Res. 2021;23(1):152-60. https://doi.org/10.1093/ntr/ntaa011 PMid:31943074 PMCid:PMC7789949
29. Stringhini S, Carmeli C, Jokela M, et al. Socioeconomic status and the 25×25 risk factors as determinants of premature mortality: A multicohort study and meta-analysis of 1·7 million men and women. Lancet. 2017;389(10075):1229-37. https://doi.org/10.1016/S0140-6736(16)32380-7 PMid:28159391

30. Bind M-A. Causal modeling in environmental health. Annu Rev Public Health. 2019;40:23-43. https://doi.org/10.1146/annurev-publhealth-040218-044048 PMid:30633715 PMCid:PMC6445691
31. Umweltbundesamt. Feinstaub [Particulate matter]. 2022. Available at: https://www.umweltbundesamt.de/themen/luft/luftschadstoffe-im-ueberblick/feinstaub#undefined (Accessed 20 January 2022).
32. Statistisches Bundesamt. Was ist der Mikrozensus [What is the Micro Census]? 2022. Available at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Haushalte-Familien/Methoden/mikrozensus.html (Accessed 20 January 2022).
33. RDC of the Federal Statistical Office and Statistical Offices of the Länder. Micro Census, years: 1999, 2003, 2005, 2009, 2013, 2017, own calculations. 2021.
34. BBSR. INKAR-Indikatoren und Karten zur Raum-und Stadtentwicklung. 2022. Available at: https://www.inkar.de/ (Accessed 20 January 2022).
35. Stuart EA, Rubin DB. Matching methods for causal inference: Designing observational studies. 2007. Available at: https://www.biostat.jhsph.edu/~estuart/StuRub_MatchingChapter_07.pdf (Accessed 20 January 2022).
36. Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. J R Stat Soc. 2008;171(Part 2):481-502. https://doi.org/10.1111/j.1467-985X.2007.00527.x
37. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit Anal. 2007;15:199-236. https://doi.org/10.1093/pan/mpl013
38. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21. https://doi.org/10.1214/09-STS313 PMid:20871802 PMCid:PMC2943670
39. Black PE. Euclidean distance. 2004. Available at: https://xlinux.nist.gov/dads/HTML/euclidndstnc.html (Accessed 20 January 2022).
40. Black PE. Manhattan distance. 2019. Available at: https://xlinux.nist.gov/dads/HTML/manhattanDistance.html (Accessed 20 January 2022).
41. Pope III, CA, Coleman N, Pond ZA, Burnett RT. Fine particulate air pollution and human mortality: 25+ years of cohort studies. Environ Res. 2020;183:108924. https://doi.org/10.1016/j.envres.2019.108924 PMid: 31831155
42. 41. Pope III, CA, Coleman N, Pond ZA, Burnett RT Corrigendum to "fine particulate air pollution and human mortality: 25+ years of cohort studies". Environ Res. 2020;191:109974. https://doi.org/10.1016/j.envres.2020.109974 PMid:33181974
43. Burnham KP, Anderson DR. Multimodel inference. Understanding AIC and BIC in model selection. Sociolo Methods Res. 2004;33(2):261-304. https://doi.org/10.1177/0049124104268644
44. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of satellite observations for long-term exposure assessment and global concentrations of fine particulate matter. Environ Health Perspect. 2015;123(2):135-43. https://doi.org/10.1289/ehp.1408646 PMid:25343779 PMCid:PMC4314252
45. Beloconi A, Chrysoulakis N, Lyapustin A, Utzinger J, Vounatsou P. Bayesian geostatistical modelling of $PM_{10}$ and $PM_{2.5}$ surface level concentrations in Europe using high-resolution satellite-derived products. Environ Int. 2018;121(part 1):57-70. https://doi.org/10.1016/j.envint.2018.08.041 PMid:30179765 PMCid:PMC6295977

# APPENDIX A

**Table A1.** OLS estimations for smoking and overweight on life expectancy with traffic PM pollution

| **Background PM** | **Dependent variable: Life expectancy** | | | |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | **(4)** |
| Smoker (share) | -10.906** (4.916) | | 9,662 (16.663) | |
| Overweight (share) | | -16.069*** (3.171) | | -13.647* (7.392) |
| $PM_{10}$ | -0.016 (0.046) | -0.150*** (0.049) | | |
| $PM_{2.5}$ | | | 0.342 (0.272) | -0.074 (0.224) |
| High school diploma | 0.019*** (0.005) | 0.004 (0.006) | 0.016 (0.010) | -0.005 (0.010) |
| Unemployment per 1,000 | -0.021*** (0.002) | -0.022*** (0.003) | -0.027*** (0.005) | -0.027*** (0.006) |
| Income (1,000 Euro) | 0.0001*** (0.00003) | 0.0001*** (0.00003) | 0.00003 (0.00004) | 0.00003 (0.00003) |
| Urban | 0.249** (0.114) | 0.100 (0.134) | 0.476*** (0.170) | 0.275 (0.201) |
| East Germany | 0.617*** (0.141) | 0.953*** (0.139) | 0.235 (0.283) | 0.683*** (0.265) |
| Smoker · $PM_{10}$ | 0.110 (0.169) | | | |
| Overweight · $PM_{10}$ | | 0.390*** (0.118) | | |
| Trend | 62.466*** (6.748) | 81.307*** (8.771) | | |
| $Trend^2$ | -40.977*** (2.975) | -40.209*** (3.200) | | |
| Smoker · $PM_{2.5}$ | | | -1,094 (1.183) | |
| Overweight · $PM_{2.5}$ | | | | 0.369 (0.523) |
| Constant | 80.207*** (1.542) | 84.736*** (1.604) | 77.064*** (4.117) | 86.112*** (3.340) |
| Observations | 423 | 423 | 70 | 70 |
| Adjusted $R^2$ | 0.768 | 0.773 | 0.570 | 0.650 |
| F statistic | 140.748*** | 144.682*** | 12.438*** | 17.021*** |
| AIC | 787.974 | 778.926 | 137.442 | 123.038 |

Note. *p<0.1; **p<0.05; ***p<0.01; & Source: Own calculations

In this appendix, OLS estimation for the relationship of smoking and overweight on life expectancy and the effects of traffic $PM_{10}$ and $PM_{2.5}$ air pollution are presented. The considerably smaller number of observations for $PM_{2.5}$ concentration implies that the PM2.5 estimations are statistically insignificant at the 5% error level. Therefore, estimations (3) and (4) are not further considered. Estimations (1) and (2) concern traffic $PM_{10}$ air pollution. For the share of smokers, as well as the share of overweight people, the negative effects on life expectancy prevail. However, for the smoker estimation, $PM_{10}$, as well as the interaction smokers · $PM_{10}$, are no longer statistically significant. By contrast, these variables remain statistically significant for the overweight estimation (2), with the same signs as in the main text with background $PM_{10}$.

**Figure A1** and **Figure A2** show the moderator effect of traffic PM. Since the relevant moderator effects of traffic $PM_{2.5}$ are statistically insignificant, they are not considered here (see the lower parts of **Figure A1** and **Figure A2**). The same holds for the moderator effect of traffic PM10 for smokers and life expectancy in the upper part of Figure A1. Only the traffic $PM_{10}$ moderator effect in the upper part of **Figure A2** is relevant because of its statistical significance in estimation (2) of **Table A1**. This diagram shows that the slope of the regression line concerning the effect of overweight on life expectancy decreases (increases) as the traffic $PM_{10}$ concentration increases (decreases) by one standard deviation.

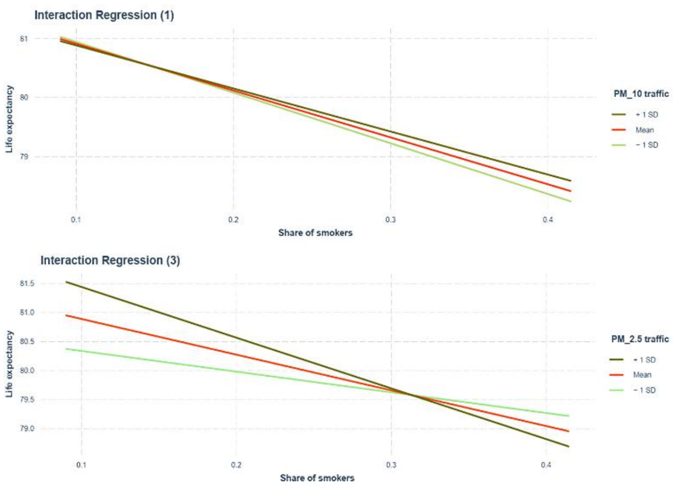


**Figure A1.** Moderator effect of traffic PM on the relationship between the share of smokers and life expectancy (Source: Own depiction with software R)

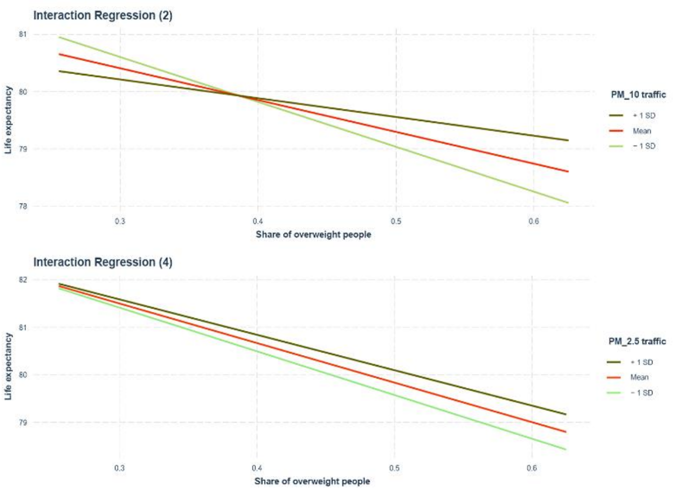


**Figure A2.** Moderator effect of traffic PM on the relationship between the share of overweight people and life expectancy (Source: Own depiction with software R)