UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Master Thesis

Faculty of Science

University of Bern

Impact of Air Pollution on Greyhound Racing Performance

handed in by

Peter Weibel

May 22, 2024

Supervisor: Prof. Dr. Eric Strobl

Advisor: Joel Hüsler

Oeschger Centre for Climate Change Research,
University of Bern

Abstract

This paper provides first evidence on the short-term effect of air pollution on the physical performance of racing greyhounds using linear regression with multiple layers of fixed effects. Based on a sample of over 40'000 individual dogs competing in 13 stadiums in Ireland between 2013 and 2020, the results suggest that increasing O_3 levels decrease race performance. Additionally, by testing for non-linearities, this analysis shows that the effect of each additional unit of O_3 decreases the race performance of greyhounds exponentially with increasing O_3 concentration. For PM_{10} and NO_2, no unambiguous effect can be confirmed reliably. Comparably low pollutant concentration levels in Ireland and the use of reanalysis data are suspected to partially drive this lack of significant results for PM_{10} and NO_2 as the results of similar studies suggest the existence of corresponding effects.

Contents

1 Introduction **3**

2 Data **7**

2.1 Greyhound Racing Data . . . 7

2.2 Air Pollution Data . . . 8

2.3 Weather Data . . . 8

2.4 Final Processing Steps . . . 10

2.5 Summary Statistics . . . 10

3 Methods **15**

3.1 Initial Situation and Model Choice . . . 15

3.2 Empirical Model . . . 15

4 Results **17**

4.1 Linear Regressions . . . 17

4.2 Non-linear Regressions . . . 20

5 Discussion **23**

5.1 Interpretation of the Results . . . 23

5.2 Reliability Assessment . . . 25

6 Conclusion **28**

A Appendix **30**

1 Introduction

As of today, an extensively growing body of literature leaves little room for doubts that exposure to air pollution negatively affects humans in different ways. Within this body, health-related effects or air pollution might be the most prominent research strand. Mostly entering the human via inhalation and ingestion, air pollutants have negative effects on both the respiratory and the cardiovascular system and lead to increased mortality and morbidity rates [Anderson et al., 2012; Kampa and Castanas, 2008].

Besides these health-centered effects which are being examined since the early 20$^{\text{th}}$ century [Pincus and Stern, 1937], more recent researchers started to shift their focus towards other effects of air pollution. Aguilar-Gomez et al. [2022] identify non-health effects[1] of air pollution by reviewing empirical studies and show that modest levels of air pollution can already have negative impacts on productivity, cognitive performance and decision making. Interested in the effect of air pollution on short-run productivity for China's aggregate manufacturing sector, Fu et al. [2021] found that a 1 $\mu g/m^3$ decrease in *particulate matter* ($PM_{2.5}$) leads to an increase in productivity by 0.82%. Zivin and Neidell [2012] produced similar results on a smaller scale based on a panel dataset on agricultural worker productivity from a large farm in California. Gatto et al. [2014] focused on the effects on cognitive functions using an existing dataset from clinical trials in California and combining it with corresponding air pollution data. With their regression model, they find significant evidence that higher exposure to *ozone* (O_3), $PM_{2.5}$ and *nitrogen dioxide* (NO_2) is associated with lower cognitive abilities. This relation has also been identified by Austin et al. [2019] using data from before and after the retrofitting of school bus engines across Georgia which decreased the emitted air pollutants. Their results show that both student health and academic achievement improved after the air pollution reduction. Generally speaking, the majority of research concerned with the effects of air pollution on humans is based on long-term effects which occur when a person lives in a highly polluted environment over a longer period of time. However, there is also scientific evidence that short-term exposure to air pollution affects humans in both health- and performance-related terms [Beavan et al., 2023; Le Tertre et al., 2002; Shehab and Pope, 2019].

Keeping the information from the previous section in mind, it is time to move to a rather specific research strand which is concerned with the impacts of short-term exposure to air pollution on physical activity. A large part of the available papers in this sector share a similar structure, usually based on an already existing dataset which is joined by air pollution data and additional case-specific data - a *natural experiment* setting [de Vocht et al., 2021]. Using regression analysis methods, the primary goal is to estimate the impact of air pollution on the case-specific dependent variable measuring some sort of physical activity. Guo and Fu [2019] used a large data set based on marathon races in China and found that runners need more time to finish the same race if the air pollution is higher. Two years before, Lichter et al. [2017] found similar results by calculating the effect of air pollution on the individual productivity of professional soccer players in Germany. The productivity was measured by the total number of passes per match, and the results confirmed that this number decreases with increasing air pollution at already moderate levels. Mullins [2018] also fits in with his analysis of intercollegiate track and field athletes, re-

[1] Effects of air pollution exposure which are not directly affecting parts of the human body.

porting a 0.39% performance reduction in endurance events for each 10 ppb increase in contemporaneous O_3.

Besides these reports centered around the effect of air pollution on human physical performance, there is also a small branch of research considered with the effect of air pollution on the physical performance of animals. Presently, all available research papers within this branch are built upon horse racing data. Gates [2007] used data from major United States horse racetracks over a 35-year time period to estimate the effect of air pollution on finish time. She found no significant effect for PM and only small negative effects for O_3 and NO_2. To explain the lack of convincing findings, the author mentions the limited size of the dataset (675 races in total), missing variables in the model and little variance in the air pollution exposure as potential shortcomings. Following Araneda [2022], physical performance is affected by intrinsic and extrinsic factors. While genetic load, degree of training, age or psychological factors belong to the intrinsic factors, extrinsic factors correspond to the real-world conditions of an event such as race distance, track condition, wind speed or air pollution[2]. An optimal study design to measure the effect of air pollution on physical performance would require athletes endowed with similar intrinsic factors performing at their best in a large number of similar events under varying air pollution conditions - a setting which happens to suit the horse racing case: The horses are bred, selected and trained exclusively for horse racing and compete regularly in similar extrinsic conditions. Therefore, the results should display an accurate estimate of the effect of air pollution on physical performance. The author employed a combination of multiple statistical tests to obtain correlation coefficients between the finish times of horse races held in Santiago, Chile, and local air pollution levels. In two consecutive research papers, higher air pollution levels were found to be correlated with slower finish times [Araneda, 2022; Araneda and Cavada, 2022].

Out of the numerous animal species which compete in races around the world, horse racing and greyhound racing are by far the most well known. While horse races are being held in numerous countries across the world, greyhound races are almost exclusively held in the United States, the United Kingdom, Ireland and Australia [GREY2K, 2024; IFHA, 2024]. In this paper, greyhounds are used as research subjects to further expand the scope of research regarding the impact of air pollution on animal athletic performance. More specifically, data from Irish greyhound races is being used to estimate the effect of three air pollutants (PM_{10}, O_3 and NO_2) on greyhound racing performance. Like horses, greyhounds are being specifically bred, selected and trained for racing and compete regularly in similar real-world conditions. Therefore, greyhound racing happens to be equally suited for this research purpose.

At this point, one might be curious about the topic of this analysis: What is the use of estimating the effect of air pollution on the performance of racing greyhounds? And who could be interested in the results? Besides expanding the scope of scientific evidence in a hardly researched field, the most important justification factor for this analysis is the well-being of the greyhounds. As mentioned earlier, air pollution affects humans through short-term exposure. Associated effects include decreased cogni-

[2] The mentioned elements are not intended to form an exhaustive list; numerous intrinsic and extrinsic factors have case-specific impacts on physical performance.

tive and physical performance. While physical performance generally improves human health whereas air pollution affects it negatively, research has not yet found solid evidence at which point the positive effect of physical performance outweighs the negative effect of air pollution or vice versa [Tainio et al., 2021]. What is clear however is that short-term exposure to air pollution both reduces the physical performance of humans and has negative effects on their health. In the case of physical performance, there is even evidence of exponentially increasing negative effects of the air pollution concentration, meaning that performance is impaired at an increasing rate for each additional unit of pollution [Lichter et al., 2017]. But the increased pollutant uptake is not only based on the pollution concentration, but also on the respiratory system: Air uptake (tidal volume, breath rate), lung perfusion, the metabolic rate and the breathing pathway (nasal or oral) affect the pollution uptake [Bigazzi and Figliozzi, 2014]. As greyhounds have a structurally similar respiratory system to humans, these factors apply to them likewise. While running on the race track, a greyhound delivers its peak physical performance for about 30 seconds. During this time, its breath rate will be at its peak and the breathing takes place during the mouth, both increasing factors for the pollutant uptake. Nevertheless, the most important factor might be the lung perfusion: Greyhounds differ physically from both humans and horses. But it is no coincidence that both greyhounds and horses are able to run really fast. Both species' hearts are much larger in comparison to their entire body mass (1-2%) as the human heart (0.5%), enabling them to have a quicker *oxygen* (O_2) transportation system which allows for a O_2 utilization three times as high as found in elite human athletes[3]. This benefit with respect to their peak physical performance on the other hand allows them to take up more pollutants and, in conjunction with the other factors, potentially exposes them to extraordinarily high health risks in comparison to human athletes. In this context, the results of this analysis could prove interesting in the discussion on animal welfare in greyhound racing.

Another aspect which has not attracted a lot of research interest yet is the relation between mental state and physical performance and its possible importance in the context of air pollution. In recent studies, researchers have found that psychological factors affect athletes performance in various ways [Beavan et al., 2023; Behm and Carter, 2021; Marcora et al., 2009; Sarkar and Fletcher, 2013; Van Cutsem et al., 2017]. As long as these psychological factors are distributed independently from the air pollution concentration among the athletes, they are only introducing noise which usually can be handled in a regression analysis. However, athletes are well aware of the environment they are performing in and thus it seems plausible that by knowing they are going to perform in a more polluted environment, they reduce their performance level unconsciously[4]. In this case, a fraction of the psychological factors affecting their performance is not independent from air pollution anymore and by estimating the effect of air pollution on performance, the physical and psychological effect are not distinguishable anymore. In contrast, the mental state of greyhounds is unlikely to be affected by the air pollution concentration as they are not able to receive the necessary information. Therefore, it is expected that a regression analysis based on greyhounds approximates the sought-after effect more accurately than its human-based

[3] They do not only share this: Highly contractile spleens, a high muscle mass percentage of more than 50%, tolerance for body temperatures up to 42°C and lungs which are bottlenecking the O_2 transportation (while still surpassing less athletic species such as humans by far) suggest that both animal species are potentially affected by air pollutants in a similar way because of their biological similarities with respect to their O_2 transportation system [Poole and Erickson, 2011].

[4] Thinking of this as *"fear from physical damages through performing in polluted areas"*, there is scientific evidence which suggests that this fear could impair the performance of an athlete [Datcu et al., 2021; Geukes et al., 2017].

equivalent.

Lastly, this analysis is not directly suited for participants in the betting industry to improve their betting strategies by considering air pollution levels among all the other factors. As greyhounds are looked at as an entire pool of entities without differentiating between them in the result section, it is not possible to predict the performance of certain greyhounds more accurately with the information found in this analysis. As the results suggest the existence of certain effects in a general sense, these results would need to be transferred to group- or even dog-specific cases in order to potentially gain an additional prediction variable for specific racing outcomes.

Before moving further, it is necessary to make an important preliminary remark: While the respiratory systems of greyhounds and humans share a similar biological structure, their performance (e.g. gas transport rates) differs remarkably [Poole and Erickson, 2011; Tsujino et al., 2005]. This means that the results of this analysis can not directly be translated to the human case, the specific magnitudes of the estimated coefficients will only apply to racing greyhounds. However, due to general biological similarities, animal studies are traditionally utilized to detect patterns or effects which are assumed to affect humans but are not directly observable. While the magnitude of the results usually does not, both the proof of existence and the polarity of the estimated effects might be transferable to the human case [Doncheva et al., 2021; Gabryś et al., 2022; Langley et al., 2007; Wendler and Wehling, 2010].

The remainder of this paper is structured as follows. Section 2 covers the process of collecting and preparing the data used in this analysis. In Section 3, the empirical model is being laid out. The results are presented in Section 4 and subsequently discussed in Section 5. At last, Section 6 concludes by wrapping up the main findings.

2 Data

To estimate the effect of air pollution on greyhound racing performance, three main data sources are used in this analysis. Each of the following subsections provides information about the choice, source and processing of the specific data variables. Afterwards, the final data set is visualized and described from multiple angles.

2.1 Greyhound Racing Data

Besides simply being considered an entertaining sport event since almost 100 years, there is an entire industry behind greyhound racing in Ireland which clearly contributes to its popularity [Täubert et al., 2007]. As a commercial semi-state body, *Greyhound Racing Ireland* (GRI)[5] controls the racing activities and sets the regulatory framework [GRI, 2024a]. The Irish government also repeatedly granted funding for the greyhound racing industry, for example €19 millions in the past November 2023 [Oireachtas, 2023]. While private companies operate the betting offices independently, they must pay a levy to the GRI for each bet. Alone in 2022, over 14'000 races involving over 80'000 individual dogs took place and a total prize money of €8.3 million was paid out [RCÉ, 2023].

On the result page of the official GRI website, it is possible to access race results ranging from the present time back to the 1960s. From 1980 onward, almost every race outcome is available[6]. Using the website interface, race results can be filtered by time period and stadium[7]. With thousands of races per year, a massive amount of data is theoretically retrievable. For the present analysis, the focus lies on the effect of air pollution on greyhound racing performance which is not necessarily depending on a data set covering multiple decades. Considering the availability of the other data sources[8], the research period for this analysis is defined as the interval from 2013 to 2020. Using a custom Python-based web scraper, all race results from the GRI website within the research period were initially saved to a separate data frame for each year. Each row corresponds to a single race result of a contesting dog in a specific race. The columns contain selected informations from the GRI race result tables: *year, date, stadium, race track length, track condition, dog name, sire name*[9]*, dam name*[10]*, weight, finish rank* and *finish time*. Observations with missing values in at least one column are being dropped. In a next step, air pollution and weather data is being combined with the race data.

[5] Also known as *Rásaíocht Con Éireann* (RCÉ).
[6] While the website does not claim any sort of completeness, random sampling suggests that not many race results should be missing since 1980.
[7] Stadiums are located in the Republic of Ireland as well as in Northern Ireland.
[8] More information about this in the following subsections.
[9] Name of the father.
[10] Name of the mother.

2.2 Air Pollution Data

The Irish *Environmental Protection Agency* (EPA) maintains an air pollution monitoring network across Ireland. The monitoring stations vary in what they measure: Some only measure PM, others also report O_3, NO_2 or other agents. Hourly readings are available on their website. On the SAFER data website, the EPA gives access to various data sources up to 2021 including various air pollutants. Unfortunately, the data availability for the research period of this analysis is rather bad. The Irish government has just recently started to install a state-of-the-art air quality monitoring network: While there were 30 measuring stations in 2016, this number increased to 107 in 2023 and soon, the monitoring network will reach its anticipated goal of 114 stations [EPA, 2017, 2023]. For this analysis, the limited number of stations covering the entire research period and reporting hourly measurements of at least two different air pollutants was no viable option. Thus, an alternative was chosen to account for air pollution: A reanalysis data set from the *Copernicus Atmosphere Monitoring Service* (CAMS) which is part of the Copernicus program of the European Union. Based on the reanalysis data availability, hourly model ensemble data for PM_{10}, O_3 and NO_2 from 2013 to 2020 is used to account for air pollution in this analysis. The grid resolution is 0.1° x 0.1° (which translates to about 7 km in Ireland) and the model ensemble is calculated based on 11 numerical air quality models [CAMS, 2022]. The data can be retrieved as a .csv file using an API; however, it is necessary to register to the *Atmosphere Data Store* (ADS) first. Due to its size, the data had to be requested month by month.

For further processing, Python was used. First, for each year, the monthly data sets were recomposed to cover the entire year. As almost all greyhound races take place around 8 p.m., only air pollution reanalysis data referring to 8 p.m. was kept in the data set [GRI, 2024b]. Then, inverse-distance weighted means were calculated for each pollutant at every stadium using a 10 km radius. As a result, for each stadium there are daily values for PM_{10}, O_3 and NO_2 which are calculated from all grid points within a 10 km radius around the stadium. Using a power factor of 2, grid points exponentially decrease in weight with increasing distance from the stadium in comparison to relatively closer grid points. Finally, the weighted mean air pollution values were being added to the race data set based on the specific date and stadium indication of each observation.

2.3 Weather Data

Thinking about control variables in a regression model aiming at estimating the effect of air pollutants on greyhound racing performance, meteorological variables are among the first to pop up. Numerous reports confirm the dependency of air pollution on meteorological conditions such as temperature, precipitation and wind [Elminir, 2005; Kayes et al., 2019; Liu et al., 2020]. Additionally, meteorological control variables are being included in the majority of reports with a study design similar to this analysis.

From the Irish meteorological service (also known as *MET eiréann*), hourly readings from past years are available using their website interface. The start of the time series is indicated for every station and it is possible to select the variables which should be downloaded. However, the interface only allows

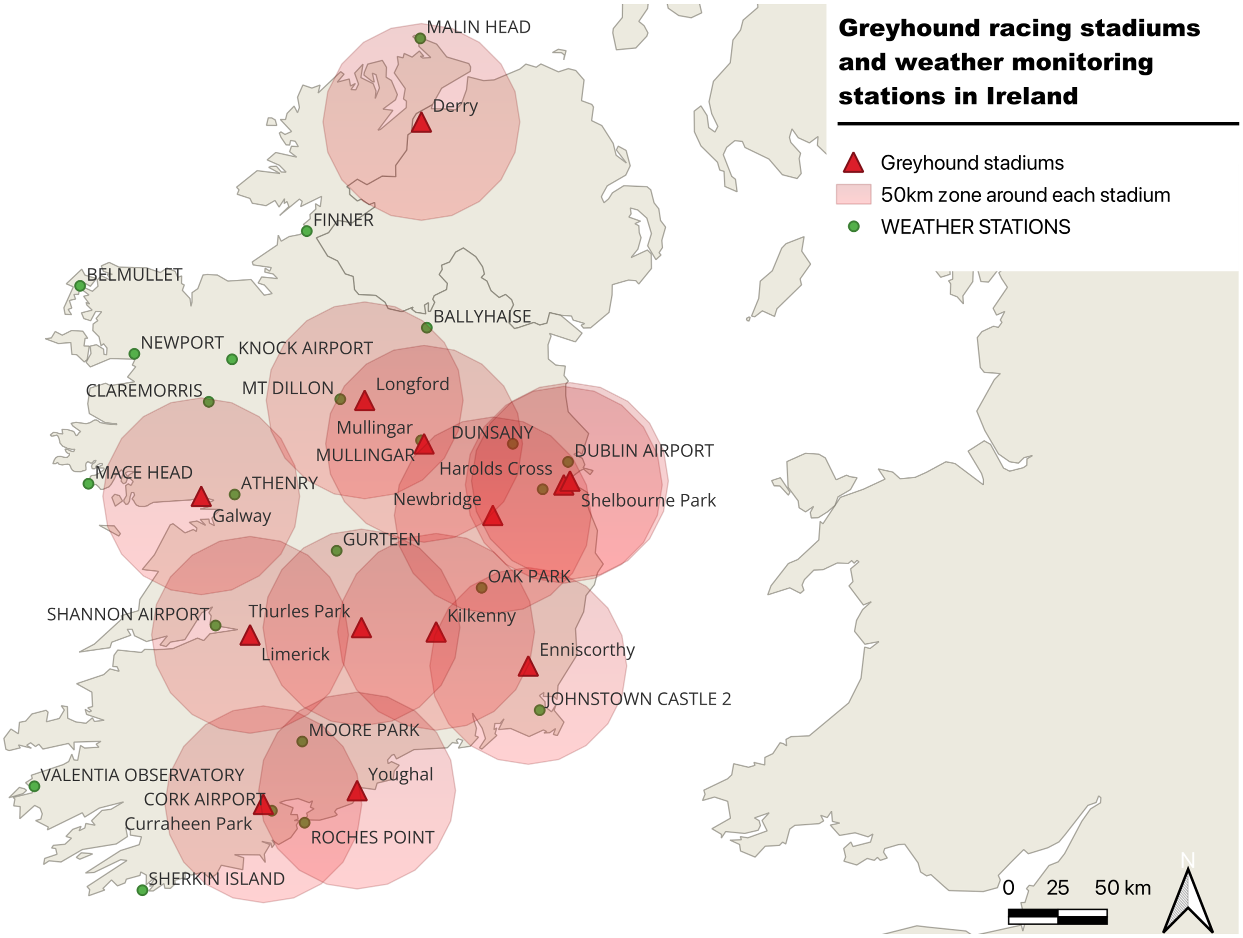

Figure 1. Overview of all greyhound racing stadiums and weather stations included in this analysis. For each stadium, all weather stations within its circular red area have been included to calculate the stadium-specific weather parameters. The exact coordinates for each stadium can be found in the Appendix (Table 6).

downloading data from one station at a time.

For the first processing step, QGIS was used. After creating a layer containing the locations of the weather stations which cover the entire study period, the coordinates of all unique greyhound racing stadiums in the data set from Section 2.1 were obtained using Google Maps, saved as a new layer and exported as .csv file. Figure 1 shows the locations of the stadiums and weather stations. Subsequently, for each stadium, the distances to all surrounding weather stations were calculated and exported as .csv file.

The following steps were carried out using Python and run for one year at a time. First, selected variables of the individual weather station data sets were imported: *date, temperature, precipitation, relative humidity, wind speed* and *predominant wind direction*. Observations with missing values were dropped. Similar to Section 2.2, only observations within the study period reporting at 8 p.m. were kept in the data set. Then, the stadium coordinate and weather station distance data sets were imported. All observations in the weather station distance data set exceeding 50 km were dropped such that only distances between weather stations and stadiums smaller than 50 km remain. Figure 1 visualizes the

50 km radius areas around each stadium. Based on this information and again similar to Section 2.2, inverse-distance weighted means were calculated for each meteorological variable at every stadium and added to the race data set based on the specific date and stadium indication of each observation; for each stadium, only data from weather stations within the 50 km radius was included. The choice of a 50 km radius reflects a compromise between minimizing the radius and simultaneously maximizing the number of weather stations within the radius area around each stadium. Also, Lichter et al. [2017] chose a comparable radius in their analysis.

2.4 Final Processing Steps

The yearly race data sets including air pollution and weather variables were first being merged to one data set covering the entire research period from 2013 to 2020. Then, observations with empty meteorological variable column values were dropped. This eliminates race results from stadiums which have no weather stations within a radius of 50 km. Further, observations with clearly identifiable outlier values were dropped[11].

Finally, four subsets from the original data set were created based on the race track length. In the original data set, there is a wide variety of race track lengths. However, around 70% (about 380'000) of the observations are in the Flat 525 category, which indicates that the races took place on 525 yards (which is equal to about 480 m) long race tracks [Towcester Racecourse, 2024]. Thus, all observations on this race track length form the main data set used in this analysis. Besides, subsets for race track lengths 330 yards (300 m), 350 yards (320 m) and 550 yards (500 m) have also been created. Each of these subsets includes around 5% of the total observation number (between 24'000 and 32'000). By re-running the calculations for each of them, it is possible to compare the results from the main data set to the secondary subsets. To some extent, this practice serves as a reliability test for the main results.

2.5 Summary Statistics

The main data set covers the research period from 2013 to 2020 and contains 376'523 race observations from 41'275 individual dogs. The races took place in 13 stadiums across the Irish island. On average, each dog participated in 9 races within the data set.

Table 1 reports the summary statistics for the variables used in this analysis. Clearing the 525 yard distance in the present Flat 525 race category takes the dogs 29.67 s on average; with the fastest time being 27.7 s, the margins seem rather small. At race day, the dogs weigh between 19.96 and 46.27 kg and on average 30.45 kg. The track condition value[12] lies between -1.3 and 0.5 with a mean of -0.15.

[11] A small number of observations with unrealistically high finish time and/or weight values.

[12] As race track conditions may vary both across stadiums and across time, the GRI regularly conducts an assessment of their state and assigns a track condition value to each stadium. As an example, a track condition value of -0.80 means that from the finish time of a dog, 0.8 s should be subtracted when comparing it to a finish time from a stadium with a track condition value of 0.

Table 1. Summary Statistics

Variable	Mean	St. Deviation	Median	Minimum	Maximum	Obs.
Finish Time (s)	29.67	0.52	29.64	27.70	35.52	376523
PM_{10} (µg/m^3)	10.54	6.09	9.23	0.63	60.90	376523
O_3 (µg/m^3)	55.30	17.20	55.37	0.73	124.79	376523
NO_2 (µg/m^3)	6.35	7.27	3.57	0.31	69.97	376523
Weight (kg)	30.45	3.25	30.39	19.96	46.27	376523
Track Condition (Value)	-0.15	0.25	-0.10	-1.30	0.50	376523
Temperature (°C)	10.58	4.28	10.54	-0.63	22.53	376523
Precipitation (mm)	0.12	0.43	0.00	0.00	5.37	376523
Relative Humidity (%)	83.54	9.29	84.58	43.47	100.00	376523
Wind Speed (kt)	8.52	4.27	7.68	1.32	29.32	376523
Predominant Wind Direction (°)	211.13	76.22	214.56	10.00	353.19	376523

Note: The descriptive statistics shown in this table refer to the main data set (only observations from the Flat 525 racetrack category are included). Detailed information about the data sources can be found in Section 2.

As for the air pollutants, the average PM_{10} concentration at race time is 10.54 µg/m^3. Minimum and maximum values are 0.63 and 60.9 µg/m^3. O_3 concentrations lie between 0.73 and 124.79 µg/m^3, the mean value is 55.3 µg/m^3. And finally, the average NO_2 concentration is 6.35 µg/m^3 with minimum and maximum values of 0.31 and 69.97 µg/m^3. Due to the influence of the Atlantic ocean, the climate in Ireland is known for its absence of temperature extremes despite its comparably high latitude [MET, 2024]. Therefore, the relatively small temperature range from -0.63 to 22.53 °C in the data with a mean of 10.58 °C is not surprising, keeping in mind that only values at 8 p.m. are included. Further, hourly precipitation lies between 0 and 5.37 mm with an average of 0.12 mm. Relative humidity ranges from 43.47% to 100% and its mean value is 83.54%. Wind speed varies between 1.32 and 29.32 kt and the predominant wind direction is south-west which coincides with the long-term weather characteristics of Ireland.

The summary statistics help to get an initial idea of the scope of the variables. However, a lot of potentially relevant information is not displayed. While the control variables do not need a closer examination, it is necessary to look at the main variables in more detail to recognize the different patterns in their variation. On the left side of Figure 2, the distribution of the observed finish times is shown. The values are approximately normally distributed around the mean value of 29.67 s and only a few observations appear to extend the upper tail further than the lower tail. The right side figure shows the distribution of the number of observations per dog. The number of dogs is exponentially decreasing as the number of observations in the data increases. This indicates that a large part of the dogs in the data set only raced a few times. However, due to the large size of the data set, there is still a substantial number of dogs with higher observation counts. Thus, this should not be a major concern.

As for the pollutants, it is well known that their composition typically changes throughout the year. Also, their spatial distribution tends to vary substantially within short distances, based on factors such as wind or the presence of local pollution sources [Bodor et al., 2020; Chen et al., 2017]. Based on a design from Lichter et al. [2017], Figure 3 visualizes both the spatial and temporal variation of PM_{10}, O_3 and NO_2 in the data set. The first row shows that the average PM_{10} concentrations seem to be the

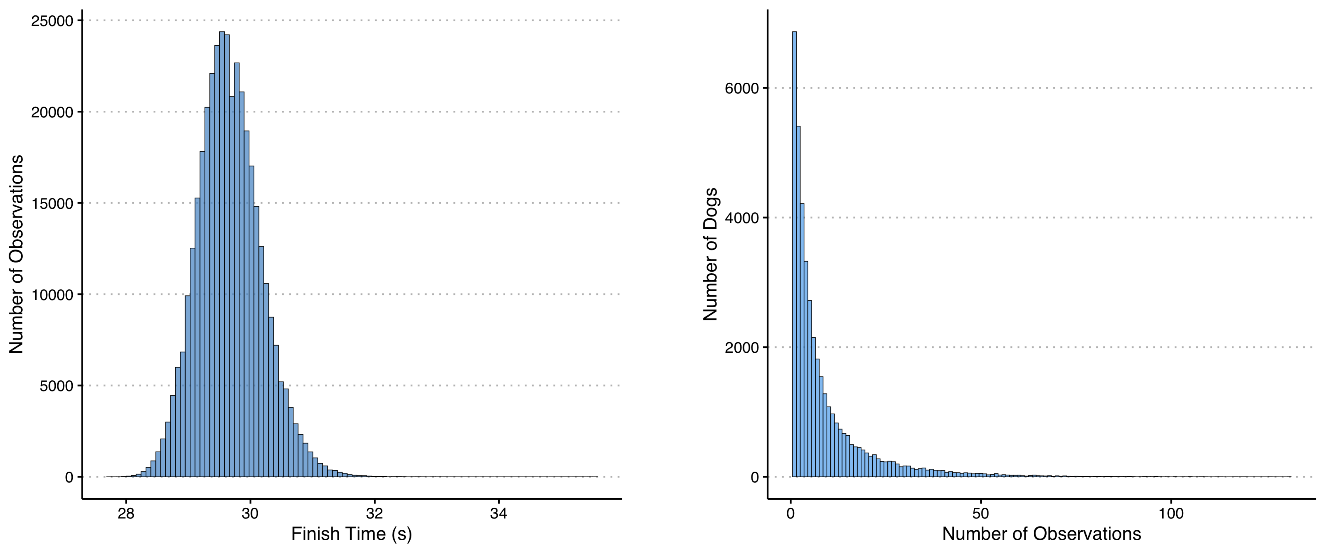

Figure 2. *Left:* Distribution of the observed finish times (n = 376'523). The range from minimal to maximal observed finish time value is divided into 100 bins and the finish time values are grouped accordingly. *Right:* Distribution of the number of observations in the data per dog. Each bin corresponds to an integer, starting at 1 and increasing step wise by 1 until the maximum value of 131 is reached.

lowest in the summer months and peaks in spring. For HRX and SPK, the average PM_{10} concentration is slightly higher than for the other stadiums. Altogether, there is both temporal and spatial variation in the concentration of PM_{10}. The magnitude, however, is really small and the PM_{10} levels are on average distinctively below the annual mean concentration limit of 40 µg/m^3 set by the European Union which is only surpassed by a few observations. Moving on to the second row, O_3 offers the largest seasonal variance of all three pollutants with a similar pattern as in the PM_{10} case: Peak concentrations in spring and then receding concentrations in the following months. In contrast to PM_{10}, the O_3 concentration peak is slightly delayed in April and the following recession extends until November. Interestingly, the O_3 concentration at the stadiums seems to be inversely correlated to the PM_{10} and NO_2 concentrations; both HRX and SPK have the lowest O_3 values while the other stadiums show little variance. While the O_3 concentration offers greater variation than the two other pollutants, the average concentration level also stays below the annual mean concentration limit of 120 µg/m^3 with only a single value exceeding this threshold. Lastly, the third row shows that the average NO_2 level is very low; however, there are more observations exceeding the annual mean concentration limit of 40 µg/m^3 than in the other two cases. While the NO_2 concentration minimum is, as in the PM_{10} case, in summer, the peak is not in spring but in early winter. The concentrations at the stadiums are similar to the PM_{10} case, but the peak concentration stadiums HRX and SPK stand out more clearly.

The following Section 3 introduces the model used to estimate the effect of the different air pollutants on finish time. Before heading there, Figure 4 visualizes the distribution of the observed PM_{10}, O_3 and NO_2 concentrations based on finish times. The observations have been grouped using a hexagonal raster which allows to display not only the distribution but also the density of the data and helps to get a first

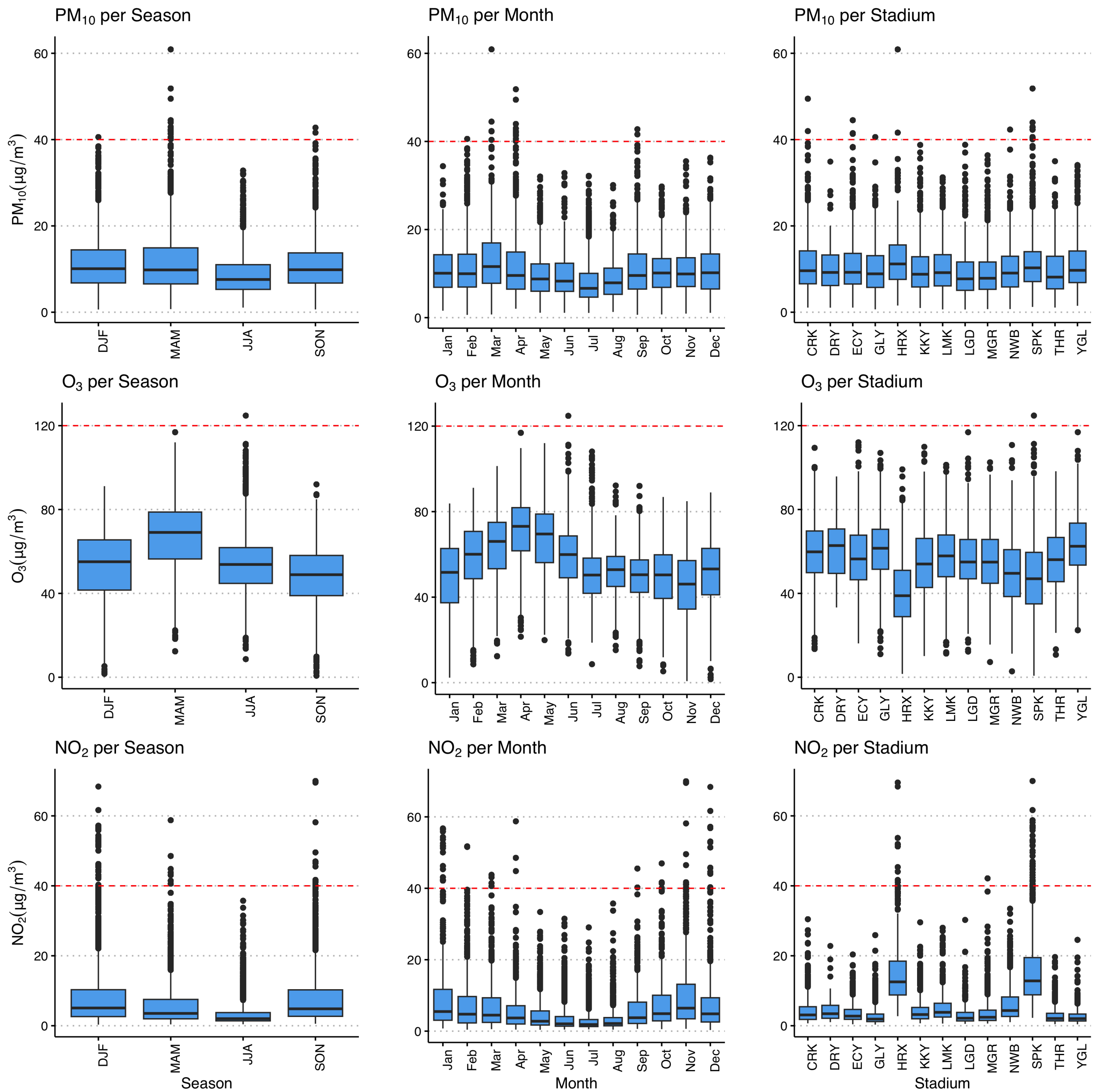

Figure 3. Conventional box-whisker plots for PM_{10}, O_3 and NO_2 concentrations per season, month and stadium (n = 376'523). Seasons are defined as following: DJF = Dec, Jan, Feb; MAM = Mar, Apr, May; JJA = Jun, Jul, Aug; SON = Sep, Oct, Nov. Full stadium names can be found in the Appendix (Table 6). The dashed red line represents the annual mean concentration limit of each pollutant set by the European Union [European Parliament and Council of the European Union, 2008].

impression of the relation between finish time and each pollutant variable[13]. The spread of PM_{10} suggests that the amount of observations decreases with increasing concentration and most observations lie around 5 - 20 µg/m^3. Regarding the spread, the same holds for NO_2; however, a large part of the observations lie at very low concentrations below 10 µg/m^3. For O_3, the observation number increases first until a concentration of about 50 µg/m^3 and decreases afterwards. The observations are also distributed more evenly around the center than in the other two plots. With respect to finish time, the plots do not show

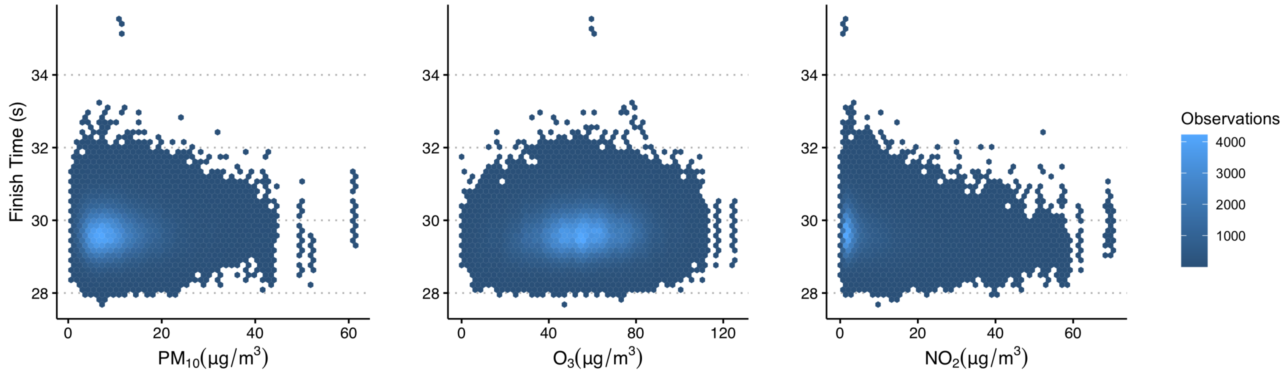

Figure 4. Distribution of the observations based on finish time and PM_{10}, O_3 and NO_2 concentration (n = 376'523). The observations have been grouped within a hexagonal raster. Based on its color, each hexagonal area contains the specified number of observations.

clear trends. The NO_2 concentration seems to be negatively correlated to finish time. However, with large parts of the observations being centered at very low concentration values, this trend might be misleading. As being elaborated in detail shortly, there are further factors which need to be considered when trying to estimate the effect of air pollution on finish times. Therefore, the main benefit from Figure 4 is the knowledge of the distribution of the three pollutants which will be relevant for the result interpretation in Section 5.

[13] The vertical alignment of observation groups is due to the pollution value matching to the races. There is always one pollution level value for each pollutant at a specific day at a specific stadium. Therefore, all observations from the same race day at the same stadium include the same pollutant concentrations.

3 Methods

3.1 Initial Situation and Model Choice

As initially stated, the goal of this analysis is to estimate the effect of air pollution (by using PM_{10}, O_3 and NO_2 concentration data) on the physical performance of greyhounds (reflected by race finish times). In line with both theoretical work and empirical case studies with similar research questions, regression analysis is the preferred method for the present analysis [Cunningham, 2021; Guo and Fu, 2019; Huntington-Klein, 2021; Lichter et al., 2017; Mullins, 2018]. Being able to disentangle the causal effect, a regression model needs to be set up such that all factors which alter the effect of the independent on the dependent variable are controlled for. Observable factors can be controlled for by including them as control variables in the model. If non-observable factors are expected to influence the estimate, several options such as *instrumental variables* (IV), *fixed effects* (FE) or *random effects* (RE) are available to control for them as well.

In the present analysis, several observable factors which are expected to have an influence on the dependent variable have been identified and are being included in the model as control variables: Weather characteristics, dog weight and race track condition. However, there are also various non-observable factors which might distort the model estimations. Adding four levels of fixed effects to the model controls for these non-observable factors. With that, the variation between the different entities/categories of each fixed effect level is being withdrawn from the model as only the variation within each entity/category remains [Huntington-Klein, 2021]. To account for time-invariant differences between dogs (e.g. different sensitivity to PM_{10}), dog fixed effects are being used. With stadium fixed effects, time-invariant differences between stadiums (e.g. average noise level through audience cheering) can be controlled for. Season fixed effects take the seasonal changes of the air pollution levels into account and year fixed effects make sure that time-specific variation is being captured (e.g. introduction of a new rule within the research period which affects the finish times of all dogs equally).

3.2 Empirical Model

Based on the scientific evidence presented in Section 1, air pollution is expected to have a negative impact on performance. As mentioned earlier, to reliably estimate the causal effect of air pollution on sport performance, several empirical challenges need to be tackled. In the present setting with greyhound racing results as the main data source and by controlling for both observable and non-observable characteristics, the estimated coefficients are expected to reflect a good approximation of the true effect of air pollution on greyhound racing performance.

To find out whether the initially stated expectation holds for greyhound racing outcomes in Ireland, the following empirical model has been constructed:

$$time_{irds} = \beta_1 PM10_{ds} + \beta_2 Ozone_{ds} + \beta_3 NO2_{ds} + X'_{irds}\gamma + \boldsymbol{W'_{ds}\delta} + R'_{ds}\mu + \alpha_i + \sigma_s + \kappa_z + \nu_y + \epsilon_{irds}, \quad (1)$$

where the finish time of dog i for race r on day d in stadium s ($time_{irds}$) is regressed on the PM_{10}, O_3 and NO_2 concentration at the corresponding day and location of the race ($PM10_{ds}$, $Ozone_{ds}$, $NO2_{ds}$). $X'_{irds}\gamma$ contains controls for individual greyhound characteristics: In this analysis, only weight is being included. To account for the influence of weather conditions on both air pollution as well as individual productivity of the greyhounds, $\boldsymbol{W'_{ds}\delta}$ includes controls for various weather variables such as mean air temperature, total precipitation, relative humidity, mean wind speed and wind direction during the hour following the approximate race start. $R'_{rds}\mu$ encloses controls for race-specific features, namely the race track condition. Finally, several fixed effects are added to control for potentially unobserved differences. α_i contains dog-specific fixed effects to control for unobserved time-invariant differences between the dogs, σ_s includes fixed effects for each stadium to account for time-invariant differences across the race track locations, κ_z features seasonal fixed effects to capture seasonal patterns within the data and ν_y takes care of time trends through yearly fixed effects. ϵ_{irds} denotes the error term.

Equation 1 is the baseline model specification. In the following Section 4, several adaptations of the baseline model will be used; however, no conceptual changes to the model will occur. Therefore, the variations of Equation 1 are not listed separately.

4 Results

4.1 Linear Regressions

Table 2 shows the results of the baseline regression model including all three pollutants, control variables and fixed effects. Columns 1 to 4 represent the stagewise addition of fixed effects as visible in the second panel. While the coefficient for PM_{10} is still highly significant in column 1, this changes subsequently when introducing further fixed effects with no significance remaining at all. In contrast, the coefficients for both O_3 and NO_2 maintain their statistical significance on the 0.01 level in all four columns and suggest that a 1 µg/m^3 increase in the concentration of O_3 (NO_2) increases the finish time by 0.0006 (0.0007) s, given that all other variables are being held constant. While these numbers are small, one needs to consider the magnitude of pollution level variability. O_3 has the potential to vary remarkably within hours at the same location: As an example, between February 21, 2024 at 2 p.m. and February 22, 2024 at 9 a.m., the O_3 concentration in Kilkenny dropped from 79 to 39 µg/m^3 as reported on the EPA

Table 2. Full Model Results with Stagewise Fixed Effects

Dependent Variable:	Finish Time (s)			
Stage:	(1)	(2)	(3)	(4)
Variables				
PM_{10} (µg/m^3)	0.0004***	9.75×10^{-5}	0.0002	0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
O_3 (µg/m^3)	0.0010***	0.0011***	0.0010***	0.0006***
	(5.81×10^{-5})	(5.73×10^{-5})	(6.16×10^{-5})	(6.2×10^{-5})
NO_2 (µg/m^3)	-0.0015***	0.0009***	0.0011***	0.0007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Weight (kg)	0.0132***	0.0133***	0.0120***	0.0149***
	(0.0014)	(0.0014)	(0.0013)	(0.0014)
Track Condition (Value)	-0.0221***	0.2920***	0.3058***	0.2184***
	(0.0066)	(0.0102)	(0.0101)	(0.0097)
Temperature (°C)	-0.0058***	-0.0058***	-0.0054***	-0.0054***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Precipitation (mm)	0.0174***	0.0205***	0.0199***	0.0210***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)
Relative Humidity (%)	-0.0002***	-0.0003***	-0.0002**	-0.0002**
	(7.89×10^{-5})	(7.79×10^{-5})	(7.96×10^{-5})	(7.94×10^{-5})
Wind Speed (kt)	0.0025***	0.0038***	0.0035***	0.0036***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Predominant Wind Direction (°)	$-5.23 \times 10^{-5***}$	$-3.75 \times 10^{-5***}$	$-4.81 \times 10^{-5***}$	$-5.18 \times 10^{-5***}$
	(9.44×10^{-6})	(9.41×10^{-6})	(9.39×10^{-6})	(9.34×10^{-6})
Fixed-effects				
Dog	Yes	Yes	Yes	Yes
Stadium		Yes	Yes	Yes
Season			Yes	Yes
Year				Yes
Fit statistics				
Observations	376,523	376,523	376,523	376,523
R^2	0.56227	0.56925	0.57087	0.57310
Within R^2	0.00991	0.01418	0.01094	0.00738

Note: Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

air quality monitoring website. This difference of 40 µg/m^3 would increase the finish time of a greyhound by 0.028 s based on the present regression model. Values of this magnitude are regularly separating the first from the second place or could decide whether a dog is able to claim the stadium race track record or not. Also, looking at stages 1 to 3, the estimated coefficients for O_3 and NO_2 are higher than in stage 4, suggesting that without year fixed effects, the effect of either of them on finish time would be even more pronounced. As for the control variables, all of them are significant (on the 0.05 or 0.01 level) throughout all stages.

By including the effects of all three pollutants simultaneously as in the baseline model in Table 2, the estimated coefficient for PM_{10} is only significant in the absence of stadium, season and year fixed effects. A possible explanation could be that PM_{10} is correlated to O_3 and/or NO_2. In fact, the correlation matrix in Table 7 in the Appendix reports a moderate correlation of 0.28 between PM_{10} and NO_2. While this value is not considered as a strong correlation value which would be an indicator for potential multi-

Table 3. Pollutant-Specific Results with Fixed Effects

Dependent Variable:	Finish Time (s)		
Independent Variable:	PM_{10}	O_3	NO_2
Variables			
PM_{10} (µg/m^3)	0.0004^{***}		
	(0.0001)		
O_3 (µg/m^3)		0.0005^{***}	
		(5.35×10^{-5})	
NO_2 (µg/m^3)			-4.62×10^{-6}
			(0.0001)
Weight (kg)	0.0149^{***}	0.0149^{***}	0.0149^{***}
	(0.0014)	(0.0014)	(0.0014)
Track Condition (Value)	0.2157^{***}	0.2168^{***}	0.2157^{***}
	(0.0097)	(0.0097)	(0.0097)
Temperature (°C)	-0.0051^{***}	-0.0055^{***}	-0.0051^{***}
	(0.0002)	(0.0002)	(0.0002)
Precipitation (mm)	0.0220^{***}	0.0210^{***}	0.0213^{***}
	(0.0017)	(0.0017)	(0.0017)
Relative Humidity (%)	-0.0004^{***}	-0.0002^{***}	-0.0004^{***}
	(7.81×10^{-5})	(7.91×10^{-5})	(7.8×10^{-5})
Wind Speed (kt)	0.0039^{***}	0.0034^{***}	0.0039^{***}
	(0.0002)	(0.0002)	(0.0002)
Predominant Wind Direction (°)	$-6.79 \times 10^{-5***}$	$-6.7 \times 10^{-5***}$	$-7.64 \times 10^{-5***}$
	(9.1×10^{-6})	(8.74×10^{-6})	(8.98×10^{-6})
Fixed-effects			
Dog	Yes	Yes	Yes
Stadium	Yes	Yes	Yes
Season	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	376,523	376,523	376,523
R^2	0.57296	0.57307	0.57295
Within R^2	0.00704	0.00730	0.00701

Note: This table shows the estimated coefficients for the regressions with only one pollutant at a time as independent variable. Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

collinearity issues, it might still be the case that through the partial correlation, the coefficient estimate for PM_{10} is being affected [Akoglu, 2018; Shrestha, 2020].

Table 3 reports the outputs of three regressions including all fixed effect levels based on a variation of the baseline model. Each variation includes either PM_{10}, O_3 or NO_2 as independent variable, everything else remains unchanged. By isolating PM_{10} as the only pollution-related variable in the regression, its estimated coefficient of 0.0004 becomes highly significant on the 0.01 level. Together with the comparably high correlation value, this leads to the suggestion that the inclusion of NO_2 interferes with the coefficient estimate of PM_{10} in Table 2. Vice versa, the estimated coefficient of NO_2 is not significant anymore. The estimated O_3 coefficient remains almost unchanged compared to Table 2, slightly decreasing to 0.0005 but remaining highly significant on the 0.01 level.

One concern regarding the prior regressions is that many of the dogs in the data set might not have

Table 4. Top Pollution Percentile Results with Fixed Effects

Dependent Variable:		Finish Time (s)	
Restriction:	$PM_{10} \geq 40$	$O_3 \geq 80$	$NO_2 \geq 40$
Variables			
PM_{10} (µg/m^3)	0.0002		
	(0.0003)		
O_3 (µg/m^3)		0.0007^{***}	
		(6.46×10^{-5})	
NO_2 (µg/m^3)			0.0005^{**}
			(0.0002)
Weight (kg)	0.0165^{***}	0.0139^{***}	0.0100^{**}
	(0.0060)	(0.0018)	(0.0044)
Track Condition (Value)	0.2136^{***}	0.1319^{***}	0.0901^{***}
	(0.0373)	(0.0130)	(0.0282)
Temperature (°C)	-0.0058^{***}	-0.0052^{***}	-0.0059^{***}
	(0.0010)	(0.0003)	(0.0007)
Precipitation (mm)	0.0164^{**}	0.0194^{***}	0.0214^{***}
	(0.0065)	(0.0020)	(0.0048)
Relative Humidity (%)	6.98×10^{-5}	-0.0003^{***}	-0.0005^{*}
	-0.0001	(9.68×10^{-5})	(0.0003)
Wind Speed (kt)	0.0035^{***}	0.0033^{***}	0.0022^{***}
	(0.0007)	(0.0002)	(0.0005)
Predominant Wind Direction (°)	2.77×10^{-5}	$-7.46 \times 10^{-5***}$	3.37×10^{-5}
	(4.26×10^{-5})	(1.09×10^{-5})	(2.79×10^{-5})
Fixed-effects			
Dog	Yes	Yes	Yes
Stadium	Yes	Yes	Yes
Season	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	18,822	220,308	37,531
R^2	0.50873	0.50815	0.53842
Within R^2	0.00777	0.00608	0.00450

Note: This table shows the estimated coefficients for the pollutant-wise regressions based on restricted data sets: For each regression, only observations from dogs which have at least one observation with a pollutant concentration value above the specified value are being included. Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

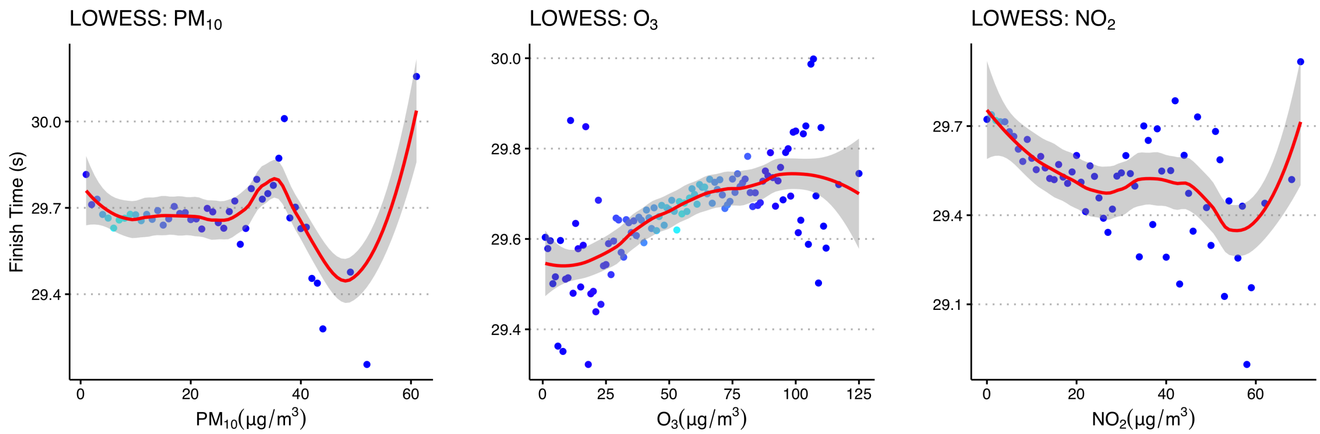

Figure 5. *Locally weighted scatterplot smoothing* (LOWESS) based plots for the distribution of finish times and PM_{10}, O_3 and NO_2. Observations from the main data set (n = 376'523) have been grouped based on their pollution levels (rounded to the next integer) such that each dot represents the mean finish times at a given pollution level. Brighter points indicate that they contain more observations than darker points. The red line represents the results from the LOWESS estimation of each pollutant on finish time without additional control variables. The grey areas denote the standard error.

a lot of pollution level variation in their observations[14]. To address this concern, a similar model variation to the one used in Table 3 is being used. However, the main data set is being filtered based on a simple restriction: Only dogs which have at least one observation with the respective pollutant concentration value above the respective EU annual mean threshold value[15] are being included in the data set[16]. By that, each regression uses a specifically filtered data set which might be more suitable to estimate the effect of pollution on finish time as there is a larger variety in pollution values for each dog compared to the unfiltered case.

Table 4 summarizes the regression results including the aforementioned restrictions. The role of PM_{10} and NO_2 has changed again back to the situation in the baseline model in Table 2 with the coefficient estimate of PM_{10} not being significant while NO_2 reaches the 0.05 significance level with a coefficient estimate of 0.0005. Once again, the O_3 coefficient estimate of 0.0007 did not change drastically and remains highly significant on the 0.01 level. In contrast to the earlier regressions, a few control variables are less significant or not significant anymore. This might be connected to the decrease in the number of observations as both PM_{10} and NO_2 show less significant control variables and simultaneously use far less observations to estimate the coefficients as visible in the bottom panel of Table 4. O_3, which still uses a really high observation number, shows no significance issues in the control variable section.

4.2 Non-linear Regressions

In the previously estimated regression models, a linear effect of the air pollutants on finish time was implicitly assumed and many of the estimated coefficients were significant. However, it is possible that a non-linear approach would fit the data better and by that lead to more accurate results [Cunningham,

[14] As shown in Figure 4, large parts of the observations are contained within narrow pollution bands, especially for NO_2.

[15] For O_3, due to the lack of observations above the EU annual mean threshold of 120 µg/m^3, a lower threshold of 80 µg/m^3 has been chosen.

[16] E.g. the threshold for PM_{10} is 40 µg/m^3; therefore, only dogs which have at least one observation with a corresponding PM_{10} value above 40 µg/m^3 are being included in the data set used for the PM_{10} regression.

2021; Huntington-Klein, 2021]. Based on *locally weighted scatterplot smoothing* (LOWESS), Figure 5 visually represents the relationship between pollutants and finish times without including any further control variables. The LOWESS regression lines for PM_{10} and NO_2 share similar characteristics on the first glance. However, as in both cases the majority of data points is concentrated at low pollution levels, it is questionable how much interpretation weight should be given to the fluctuations at higher pollution levels. The LOWESS regression line for O_3 is closer to being linear; however, the slope fluctuates as well when moving away from the main body of observations.

Altogether, for PM_{10} and NO_2, the visual assessment suggests that accounting for non-linearity in the model could improve the estimated coefficients. While this is less the case for O_3, there are small slope changes visible in the plot as well which could lead to more accurate estimates when being accounted for. To test this, the main data set is being filtered such that for each pollutant, the observations are split into 2 to 3 categories based on their pollutant level. The break points between the categories are defined based on the LOWESS regression lines from Figure 5: Each category covers the pollution level segments which follow an approximately linear trend.

The stagewise results for the regressions testing for non-linear effects of PM_{10}, O_3 and NO_2 on finish time are summarized in Table 5. For PM_{10}, the observations were split in three groups based on their pollution concentration: Less than 10, 10 to 30 and more than 30 $\mu g/m^3$. As indicated by the LOWESS regression line in Figure 5, the coefficient estimate for low PM_{10} concentrations is negative and significant across all stages. In other words, up to PM_{10} concentrations of 10 $\mu g/m^3$, an increase in PM_{10} is associated with a decrease in finish time. This relationship would be contrary to the expected outcome. In contrast, albeit mostly insignificant, the estimated coefficients for higher PM_{10} concentrations suggest a positive relationship between PM_{10} and finish time which appears to hold especially for PM_{10} values above 30 $\mu g/m^3$. In the case of O_3, a positive relationship between O_3 concentration and finish time can be seen throughout all stages (significance is given at the 0.01 level except for two coefficient estimates for O_3 concentrations below 25 $\mu g/m^3$)[17]. More important, the estimated coefficients show non-linear characteristics depending on the O_3 concentration in every regression stage: With increasing O_3 concentrations, the impact of O_3 on finish time is more pronounced. In the last stage including all fixed effects, an O_3 concentration increase of 1 $\mu g/m^3$ increases the finish time by 0.0003 s when the O_3 concentration is between 25 and 75 $\mu g/m^3$, and by 0.0005 s if it exceeds 75 $\mu g/m^3$. For NO_2, the observations are only split in two groups (NO_2 concentration below or above 20 $\mu g/m^3$) due to the small variation in the data. The results suggest a negative relationship between NO_2 and finish time for the first two stages and mixed (and insignificant) effects for the third and fourth stage. Based on the estimated coefficients, no unambiguous non-linear patterns are recognizable. Again, the limited variation in the NO_2 concentration values in the data is highly suspected to render the recognition of any non-linear pattern impossible.

Wrapping up this subsection, the regressions presented in Table 5 provide interesting knowledge regarding the actual relationship between the pollutants and the finish time of the dogs, but also solidify

[17] Based on Figure 5 and the persistent loss of significance while introducing additional fixed effects, this might be due to limited data availability as with every fixed effect layer, the data set restrictions are potentially increasing.

Table 5. Non-linear Regression Results with Fixed Effects

Dependent Variable:	Finish Time (s)			
Stage:	(1)	(2)	(3)	(4)
Regression 1: PM_{10}				
$PM_{10} \times (PM_{10} < 10)$	-0.0014***	-0.0011***	-0.0012***	-0.0015***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
$PM_{10} \times (10 \leq PM_{10} < 30)$	-0.0003*	1.23×10^{-5}	6.01×10^{-5}	-0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$PM_{10} \times (PM_{10} \geq 30)$	0.0002	0.0004**	0.0002	0.0003
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Regression 2: O_3				
$O_3 \times (O_3 < 25)$	0.0008***	0.0007***	0.0005**	-5.12×10^{-6}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$O_3 \times (25 \leq O_3 < 75)$	0.0012***	0.0008***	0.0007***	0.0003***
	(6.6×10^{-5})	(6.51×10^{-5})	(6.75×10^{-5})	(6.69×10^{-5})
$O_3 \times (O_3 \geq 75)$	0.0012***	0.0010***	0.0009***	0.0005***
	(5.47×10^{-5})	(5.4×10^{-5})	(5.74×10^{-5})	(5.68×10^{-5})
Regression 3: NO_2				
$NO_2 \times (NO_2 < 20)$	-0.0045***	-0.0009***	0.0001	1.33×10^{-5}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$NO_2 \times (NO_2 \geq 20)$	-0.0024***	-0.0004***	3.19×10^{-5}	-3.18×10^{-6}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Fixed-effects				
Dog	Yes	Yes	Yes	Yes
Stadium		Yes	Yes	Yes
Season			Yes	Yes
Year				Yes
Fit statistics				
Observations	376,523	376,523	376,523	376,523
R^2 (PM_{10})	0.56107	0.56860	0.57045	0.57301
R^2 (O_3)	0.56213	0.56923	0.57084	0.57314
R^2 (NO_2)	0.56185	0.56856	0.57038	0.57295
Within R^2 (PM_{10})	0.00719	0.01270	0.00996	0.00715
Within R^2 (O_3)	0.00958	0.01413	0.01088	0.00745
Within R^2 (NO_2)	0.00896	0.01260	0.00980	0.00701

Note: This table shows the stagewise results from the regressions testing for non-linear characteristics of the relationship between finish time and PM_{10}, O_3 and NO_2. Although not reported here, each regression includes all control variables as specified in Equation 1 (control variable coefficient estimates are unambiguously significant on the 0.01 level). Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

the role of possible shortcomings of the data set itself which might impair the reliability of this analysis. As part of the aspects of this work that require discussion, these shortcomings will be examined subsequently in Section 5.

5 Discussion

Based on the empirical model and using a data set composed of various data sources, the regression results as displayed in Section 4 provide numerical estimates of the relationship between the finish time of racing greyhounds and the concentration level of the pollutants PM_{10}, O_3 and NO_2. However, the results still need to be contextualized within similar studies and their reliability needs to be critically reviewed.

5.1 Interpretation of the Results

5.1.1 Linear Regressions

In the baseline model specification (Table 2) including all three pollutants simultaneously, the estimated coefficients are all positive and in the case of O_3 and NO_2 significant on the 0.01 level for all four fixed effect stages. By running separate regressions for each pollutant (Table 3), the results change insofar as PM_{10} is now significant instead of NO_2. This lack of consistency in the coefficient estimations of PM_{10} and NO_2 has already been mentioned in the result section and will be addressed in the following subsection. Regarding the polarity of the results, positive coefficient estimates have been anticipated based on previous studies [Araneda and Cavada, 2022; Gates, 2007; Guo and Fu, 2019; Lichter et al., 2017]. As for the magnitude of the coefficients, it is not possible to rely on external results for comparison since this analysis is the first to examine the influence of air pollution on greyhound racing performance. However, internal validation is possible to some extent. By comparing the results of the main data set (which includes only observations from the Flat 525 race category) with results from the same regressions using observations from three other race category subsets (Flat 550, Flat 330 and Flat 350), a consistency evaluation of the estimated coefficients between different subpopulations of racing greyhounds can be made. Based on the plots displaying the distribution of finish times and pollution (Figures 6, 7 and 8 in the Appendix A), the observations from the data subsets share similar distribution characteristics to the main data set (Figure 4).

The regression results for the subsets are summarized in Table 8 and 9 in the Appendix A. In Table 8, all three pollutants are included simultaneously in the regression for each subset. The estimated coefficients for O_3 are mostly similar to their equivalent based on the main data set (O_3 coefficient in Table 2) across all three subsets. For PM_{10} and NO_2, the estimated coefficients appear less consistent. While the polarity of the estimated PM_{10} coefficients remains positive as in the main data set, both positive and negative NO_2 coefficients are being estimated based on the subsets. Further, the magnitude of the results differs remarkably: As an example, an increase in the PM_{10} concentration of 1 $\mu g/m^3$ increases the finish time by 0.0019 s based on the Flat 350 subset. The same regression using the main data set estimates an increase of 0.0002 s, which is an effect almost 10 times smaller. In the pollutant-specific regressions reported in Table 9, a similar pattern is showing. The estimated coefficients for O_3 are directly comparable to their equivalents based on the main data set (Table 3) with mostly significant values between 0.0004 and 0.0007 across all subsets. The estimated PM_{10} coefficients are uniformly positive and highly significant for every subset; however, the coefficient values ranging from 0.0008 to 0.0018 including all fixed effects are 2 to 4 times higher than the value of 0.0004 based on the main data set (Table 3). And for NO_2, the estimated coefficients exhibit mixed signs again.

Comparing a part of the main results to the subset results reinforces several assumptions concerning the outcome of this analysis. It is reassuring that the general direction of the estimated coefficients matches across all four data sets. This indicates that the results are not only specific to one population sample but can be replicated using different samples of the same population[18]. On the contrary, the comparison also emphasizes potential problems emerging from the data structure. The lack of pollution level variation for PM_{10} and NO_2 already led to changing coefficient signs in the main analysis; this pattern is only confirmed with the subset regressions, especially for NO_2 which has the least variation in its concentration. Ultimately, it is important to remember that this comparison does not provide new knowledge about the setup of the model or the quality of the data as it is based on the same empirical model and the same data sources as the main analysis.

Further regressions based on restricted data subsets from the main data set were carried out with the intention to increase the variation of the pollutant concentration within each individual dog by only including dogs in the regression data set which at least contain one observation above a certain threshold value[19]. By reducing the potential impact of the lack of variation in the pollution concentration within each individual dog, the resulting coefficient estimates are expected to be positive based on empirical evidence from previous studies. Table 4 summarizes the regression outcomes and following the expectations, all three coefficient estimates are positive. While O_3 is significant on the 0.01 level and NO_2 is significant on the 0.05 level, PM_{10} does not reach significance. As both NO_2 and PM_{10} reached highly significant positive coefficient estimates in the previous regressions, it seems likely that the significance issues in these regression outcomes are driven by the decrease in observations: Introducing thresholds may have increased the *quality* of the data set, but simultaneously reduced the *quantity* drastically. From initially 376'523 observations which have been used for the previous regressions, the coefficients for PM_{10} and NO_2 were estimated using only 18'822 ($\sim$5% of total) and 37'531 ($\sim$10% of total) observations, respectively. Only for O_3, the regression was based on a comparably large data subset containing 220'308 ($\sim$60% of total) observations.

5.1.2 Non-linear Regressions

With the final regressions in Section 4.2, the possibility of non-linear relationships between the air pollutants and the finish time of the greyhounds has been explored. In previous studies, corresponding results[20] were obtained in most but not all cases [Lichter et al., 2017; Matt et al., 2016; Mullins, 2018; Zhang et al., 2022]. The implementation of non-linearity in this analysis follows Lichter et al. [2017]; however, the results as shown in Table 5 are not as clear as in their report. In the case of PM_{10}, highly

[18] The population in this case would be the entirety of racing greyhounds in Ireland and the samples would be the greyhounds within each of the different data subsets used in this analysis.

[19] Especially in the case of PM_{10} and NO_2, chances are high that an individual dog only enters the data set with observations with low pollution values and thus lacks in pollution variation which is needed to estimate the effect of air pollution on physical performance.

[20] These studies are centered on the effect of air pollution on human performance.

significant effects are only found for concentrations below 10 µg/m^3 and the negative coefficient values are not in line with the expectations. Comparing this to Lichter et al. [2017], such an opposite effect for low pollution levels has also been found in their case. Nevertheless, their results are insignificant for low pollution concentrations whereas for higher pollution concentrations, the estimated coefficients are again in line with the expectations and suggest a non-linear effect of increasing pollution concentrations levels on physical performance. In this analysis, the results for higher PM_{10} concentrations are almost exclusively insignificant which prevents the confirmation of non-linear effects of PM_{10} on the physical performance of greyhounds. For NO_2, a confirmation of non-linear effects is also not possible when looking at the estimated coefficients in the later stages. By looking at the significant estimates and the distribution of the observations based on their pollutant values (Figure 5), suspicion arises that the lack in pollution concentration variation once more renders the recognition of non-linear patterns impossible. This is only supported by the fact that in the case of O_3[21], significant coefficient estimates for higher O_3 concentrations are found across all stages. Additionally, the coefficients suggest the existence of non-linear effects in the expected way: With increasing O_3 concentration, the marginal effect of each additional unit of O_3 is larger. Or in other words, the greyhounds are expected to run slower when the O_3 concentration increases from 60 to 70 µg/m^3 compared to an increase from 10 to 20 µg/m^3.

5.1.3 Impact of Outliers

By plotting the main data set pollutant distributions for PM_{10}, O_3 and NO_2, Figure 4 reveals that there are some visible outliers[22] in each subplot. In a linear regression, these outliers may have a disproportionately large influence on the slope of the regression line and thus could distort the coefficient estimates, especially if their associated dependent variable values would not be in line with the expected values based on the other observations. To check if these outlier values have a noticeable influence on the estimated coefficient, the pollutant-specific regressions have been re-estimated and for every regression, the pollutant-specific observations with outlier values have been excluded. By comparing the estimated coefficients without outliers (Table 10 in the Appendix A) to the estimated coefficients including outliers (Table 3), virtually no difference can be found. Thus, the outliers in this analysis are either 1) too few in number or 2) too close to the bulk of observations or 3) too close to the regression line to have an influence on the regression results - or any combination of them.

5.2 Reliability Assessment

5.2.1 Data

Focusing on the reliability of the data used in this analysis, there are a few important points to mention. First of all, the entire analysis is based on reanalysis air pollution data. While this specific reanalysis data set surpasses other reanalysis data sets in approximating the real world situation, it is still only a model which reaches its limits in the prediction of small-scale temporal and spatial pollution concentration variation [Zuo et al., 2023]. In addition to this, the pollution levels for PM_{10}, O_3 and NO_2 in

[21] Observations have a larger variation for O_3 as for PM_{10} and NO_2.
[22] In this case observations which are not directly connected to the large observation bulk.

Ireland are generally low[23]. While air pollution already is hard to model due to its rapidly changing spatial and temporal distribution, the generally low levels in Ireland make it even harder to capture the relevant information needed for a reliable representation of the real air pollution distribution. Thus, the reanalysis data set used in this analysis is expected to decrease in accuracy when narrowing the area or time period down to a few points of interest at a time. Unfortunately, this is exactly what this analysis seeks: Information on air pollution levels from a certain stadium at a specific hour. In the results section, the relationship between air pollution and finish time was unveiled as expected in the case of O_3 and also partially for the other pollutants, indicating that the reanalysis data was capable of capturing the real air pollution distribution at least partially. Nevertheless, the results might be even more meaningful and clear when the same analysis was carried out using real-world data, particularly in the case of PM_{10} and NO_2.

A different aspect is the role of singletons. Out of the thousands of dogs included in the main data set, a considerable fraction only appears with a single observation. Singleton observations might impair the results by underestimating the standard errors and overestimating the significance (p-value). In this analysis, singletons have not been dropped. Before applying any sort of fixed effects, they make up about 16.6% of the total number of individual entities (dogs) in the data set[24]. By introducing fixed effects, these numbers increase even more [Correia, 2015]. The effect of singletons on the results of this analysis is not further examined. Nevertheless, as a potential bias source, this topic needs to mentioned.

5.2.2 Empirical Model

By including control variables and fixed effects, the model setup in this analysis aims at isolating the effect of the three chosen air pollutants on the finish time of greyhound races. However, leaving a relevant variable uncontrolled for introduces bias which could have severe effects on the resulting coefficient estimates [Cunningham, 2021]. While the present model was constructed to the best knowledge available, there is at least one aspect which was not controlled for directly: Intertemporal effects of pollution. Mullins [2018] found that athletes training in environments with low O_3 concentrations during the week before an event are remarkably more affected by the effect of O_3 on their performance than athletes training in higher O_3 concentration environments. In this analysis, this effect is controlled for insofar as the dog fixed effects make sure that only within-group variation is used for the regressions. Thus, as long as each dog has a fixed training location and the average pollution levels at this location remain more or less constant, the results in this analysis should be unbiased by the intertemporal effects of pollution. But if the training location was to change between races or the average pollution levels at a fixed training location were to change drastically within a short timescale, the results might be biased. Under the case-specific circumstances in Irish greyhound racing, the chances of occurrence of this sort of bias are luckily quite low. Generally, most of the greyhounds are training at a fixed training location[25] and

[23] Despite being a concern for the Irish government, local air pollution levels are well below the levels found in mainland Europe [EPA, 2024].

[24] For the subsets, singleton numbers were higher. Flat 550: 38.3%, Flat 330: 39.8%, Flat 350: 39.3%.

[25] Usually at their trainers preferred training location.

in most areas in Ireland, it is unlikely that the average pollution levels change drastically between races[26].

Further, no indicator of the performance level of the greyhounds has been included in the regression models. However, such an indicator is available in the data set, namely the odds ratio. This number is based on the bidders and bookmakers expectations of the performance of each specific dog. A high value indicates a high payout for each bet on the dog in case of a victory, simultaneously meaning that this dog is not expected to win - the high potential payout serves as kind of a risk compensation. Low odd ratios subsequently indicate a higher expected winning probability [Oxford Stadium, 2024]. By re-calculating the first two regressions from the main results section, this time including the odds ratio as a control variable, the resulting coefficients remained practically unchanged (Table 11 and Table 12 in the Appendix A)[27]. Based on these results, it is not necessary to recalculate the entire empirical part of this analysis as the results do not appear to change remarkably when including or excluding the odds ratio as a control variable.

[26] Additionally, with season and year fixed effects, longer-term temporal changes in pollution levels are already controlled for.

[27] As some observations had missing or non-identifiable odd ratio entries and were dropped, the total observation number in these regressions is slightly smaller than in the main analysis (373'526 instead of 376'523).

6 Conclusion

Air pollution affects humans in various ways. Long-term exposure endangers human health by damaging the respiratory and cardiovascular system and is further known to impair cognitive performance [Anderson et al., 2012; Austin et al., 2019; Gatto et al., 2014; Kampa and Castanas, 2008]. But also short-term exposure is associated with effects such as impaired physical performance [Guo and Fu, 2019; Lichter et al., 2017; Mullins, 2018]. This analysis transfers the latter to the animal case, more specifically estimating the effect of three air pollutants (PM_{10}, O_3 and NO_2) on the finish time of racing greyhounds in Ireland. As of now, only a few comparable studies have been published which use racing horses as research subjects [Araneda, 2022; Gates, 2007]. Their results suggest either a negative effect of air pollution on physical performance or no significant effect altogether. Based on a comprehensive data set containing over 370'000 observations from more than 40'000 individual greyhounds racing in 13 stadiums between 2013 and 2020 which is paired with weather and air pollution data (Section 2), this analysis employs a linear regression model which is further extended with four layers of fixed effects to estimate the relationship between air pollution and greyhound racing performance (Section 3). In the baseline model under inclusion of all four fixed effect layers (Table 2), the estimated coefficients for O_3 (0.0006) and NO_2 (0.0007) are positive and significant; PM_{10} (0.0002) is not significant. By running separate regressions for each pollutant (Table 3), PM_{10} (0.0004) and O_3 (0.0005) are significant but NO_2 ($< |0.0001|$) is insignificant. This pattern remains throughout the further regressions: The coefficient estimates for O_3 remain strictly positive and significant while for PM_{10} and NO_2, significance and polarity change abruptly between regressions or within different stages of the same regression. At least for O_3, testing for non-linearities further reveals that the effect of each additional unit of O_3 decreases the race performance of greyhounds exponentially with increasing O_3 concentration (Table 5). The mixed results for PM_{10} and NO_2 might be partially explained by the small variation in the respective pollutant concentration as visible in Figure 4. Also, as this analysis relies on reanalysis air pollution data, this might as well have an impact on the results as the model might not be able to accurately reproduce small changes in air pollutant concentrations which is an essential condition for this analysis.

Altogether, the results of this analysis are not contrary to the expectations: In the case of O_3, solid evidence has been found that an increase in its concentration does not only reduce the physical performance of racing greyhounds but the marginal effect is exponentially intensifying with higher O_3 concentrations. And although the expected similar effects for PM_{10} and NO_2 cannot be confirmed, the coefficient estimates found in this analysis neither suggest contrasting effects; the lack of consistent significant effects only shows that based on this specific analysis, no clear relationship between either PM_{10} or NO_2 and the performance of racing greyhounds can be confirmed.

Further work on this field could resolve several uncertainties which are associated with the results of this analysis. With respect to the lack of variation in the pollutant concentrations, an analysis based on greyhound race observations from another country such as Australia would potentially offer a broader range of pollutant concentration values. And ruling out the uncertainty of the impact of reanalysis data on the results will not be an issue for further work based on the Irish case as the Irish government extended the air quality monitoring network gradually since 2017 [EPA, 2023]. Thus, in a few years from now, working with a research period from 2023 onwards allows the usage of air pollution data from over 100 measuring stations across Ireland (instead of about 30 available for the period from 2013 to 2020) which eliminates the need for reanalysis data.

A Appendix

A.1 Stadium Information

Table 6. Stadium Information

Stadium	Abbreviation	Latitude	Longitude
Curraheen Park	CRK	51.8762016558209	-8.54911570072719
Derry	DRY	54.9905919422089	-7.3364784165418
Enniscorthy	ECY	52.5007466707794	-6.57530822487508
Galway	GLY	53.2791781241726	-9.03967697283273
Harolds Cross	HRX	53.323704535055	-6.2774069540582
Kilkenny	KKY	52.6625973591212	-7.26191214297558
Limerick	LMK	52.6494164259788	-8.65849423328302
Longford	LGD	53.7216532788694	-7.79325962258396
Mullingar	MGR	53.5215534128107	-7.33923191430418
Newbridge	NWB	53.190780536834	-6.82211873888551
Shelbourne Park	SPK	53.3405403453828	-6.22971747670822
Thurles Park	THR	52.684521511899	-7.82218211450553
Youghal	YGL	51.9401200601086	-7.8572566825163

Note: This table provides both the full names as well as the abbreviations for the greyhound race stadiums featured in this analysis. Also, the exact coordinates of each stadium are presented.

A.2 Correlation Table

Table 7. Correlation Matrix

	Finish Time	PM_{10}	O_3	NO_2	Weight	Track Cond.	Temp.	Rain	Rel. Hum.	W. Speed	W. Dir.
Finish Time	1.00	-0.00	0.08	-0.13	-0.17	-0.12	-0.05	0.02	0.00	0.04	-0.00
PM_{10}	-0.00	1.00	0.05	0.28	0.00	0.08	-0.06	-0.12	-0.01	-0.09	-0.29
O_3	0.08	0.05	1.00	-0.44	-0.01	-0.02	0.05	0.07	-0.20	0.22	-0.06
NO_2	-0.13	0.28	-0.44	1.00	0.03	0.17	-0.25	-0.00	0.11	-0.22	-0.19
Weight	-0.17	0.00	-0.01	0.03	1.00	0.11	0.03	-0.00	-0.01	-0.01	-0.00
Track Cond.	-0.12	0.08	-0.02	0.17	0.11	1.00	0.01	-0.03	-0.01	-0.02	0.01
Temp.	-0.05	-0.06	0.05	-0.25	0.03	0.01	1.00	-0.01	-0.14	-0.02	0.01
Rain	0.02	-0.12	0.07	-0.00	-0.00	-0.03	-0.01	1.00	0.29	0.29	-0.14
Rel. Hum.	0.00	-0.01	-0.20	0.11	-0.01	-0.01	-0.14	0.29	1.00	0.04	-0.22
W. Speed	0.04	-0.09	0.22	-0.22	-0.01	-0.02	-0.02	0.29	0.04	1.00	0.12
W. Dir.	-0.00	-0.29	-0.06	-0.19	-0.00	0.01	0.01	-0.14	-0.22	0.12	1.00

Note: Correlation matrix of all variables which are included in the baseline regression model given by Equation 1.

A.3 Flat 550, Flat 330 and Flat 350

A.3.1 Finish Time Distribution based on PM_{10}, O_3 and NO_2 for Flat 550, 330 and 350

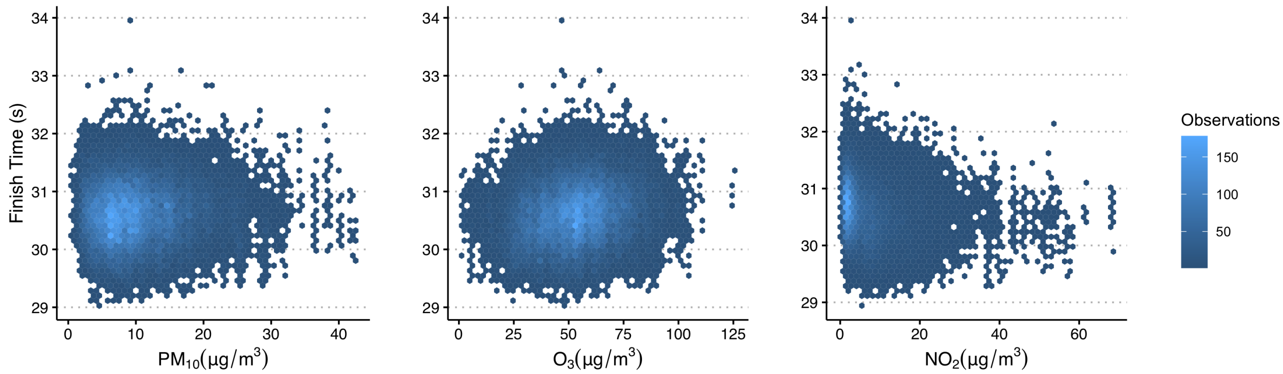

Figure 6. Distribution of the Flat 550 observations based on finish time and PM_{10}, O_3 and NO_2 concentration (n = 32'455). The observations have been grouped within a hexagonal raster. Based on its color, each hexagonal area contains the specified number of observations.

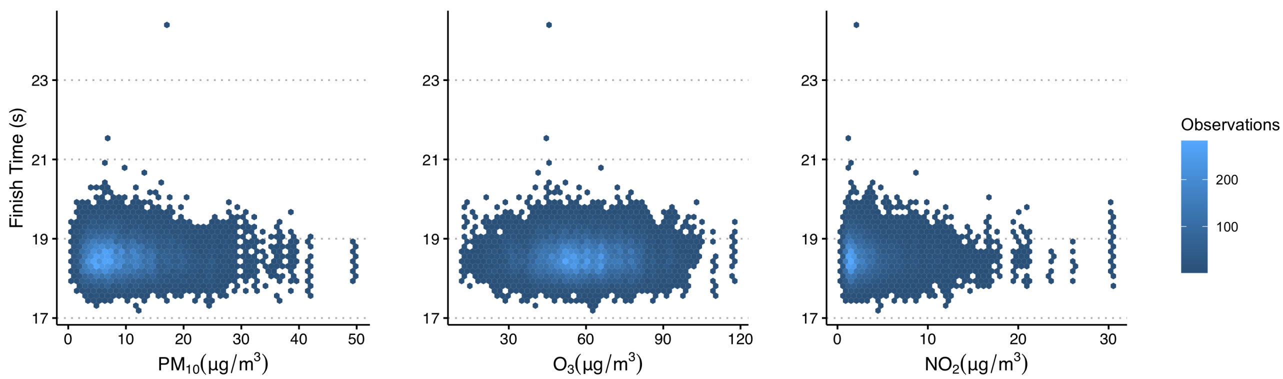

Figure 7. Distribution of the Flat 330 observations based on finish time and PM_{10}, O_3 and NO_2 concentration (n = 24'511). The observations have been grouped within a hexagonal raster. Based on its color, each hexagonal area contains the specified number of observations.

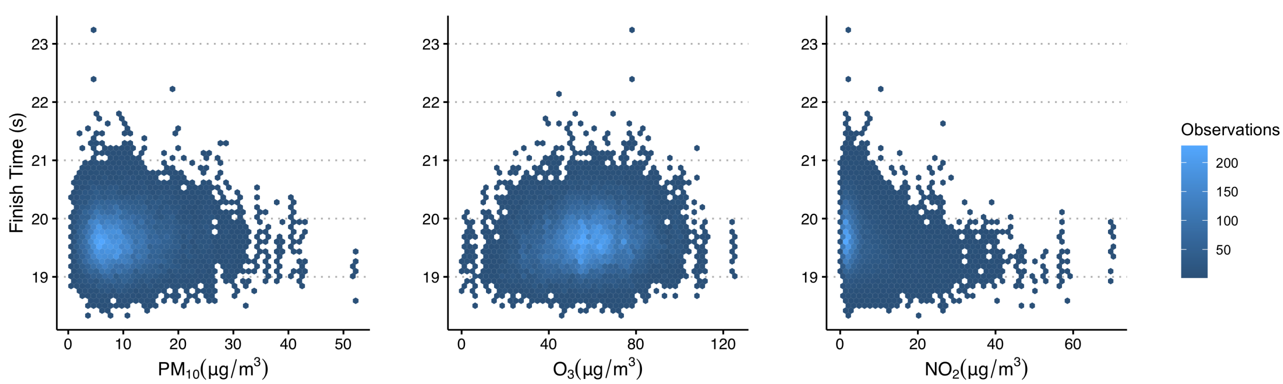

Figure 8. Distribution of the Flat 350 observations based on finish time and PM_{10}, O_3 and NO_2 concentration (n = 27'329). The observations have been grouped within a hexagonal raster. Based on its color, each hexagonal area contains the specified number of observations.

A.3.2 Selected Regression Results for Flat 550, F330 and 350

Table 8. Baseline Model Results for Flat 550, 330 and 350 with Fixed Effects

Dependent Variable:	Finish Time (s)			
Stage:	(1)	(2)	(3)	(4)
Flat 550				
PM_{10} (µg/m^3)	0.0007	4.38×10^{-5}	6.36×10^{-5}	0.0004
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
O_3 (µg/m^3)	0.0008***	0.0007***	0.0007***	6.68×10^{-5}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
NO_2 (µg/m^3)	-5.47×10^{-5}	0.0014***	0.0014***	0.0004
	(0.0004)	(0.0005)	(0.0005)	(0.0005)
Flat 330				
PM_{10} (µg/m^3)	0.0006	0.0006	0.0006	0.0005
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
O_3 (µg/m^3)	0.0004**	0.0005**	0.0007***	0.0005**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
NO_2 (µg/m^3)	0.0013	0.0014	0.0021**	0.0015
	(0.0011)	(0.0010)	(0.0011)	(0.0011)
Flat 350				
PM_{10} (µg/m^3)	0.0021***	0.0020***	0.0017***	0.0019***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
O_3 (µg/m^3)	0.0003	0.0003	0.0006***	0.0006***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
NO_2 (µg/m^3)	-0.0022***	-0.0012**	-0.0008	-0.0009
	(0.0005)	(0.0006)	(0.0006)	(0.0006)
Fixed-effects				
Dog	Yes	Yes	Yes	Yes
Stadium		Yes	Yes	Yes
Season			Yes	Yes
Year				Yes
Fit statistics				
Observations (Flat 550)	32,455	32,455	32,455	32,455
Observations (Flat 330)	24,511	24,511	24,511	24,511
Observations (Flat 350)	27,329	27,329	27,329	27,329
R^2 (Flat 550)	0.68864	0.69526	0.69629	0.69846
R^2 (Flat 330)	0.63350	0.63373	0.63432	0.63839
R^2 (Flat 350)	0.63113	0.63535	0.63714	0.63963
Within R^2 (Flat 550)	0.01648	0.02547	0.02007	0.01682
Within R^2 (Flat 330)	0.00715	0.00751	0.00500	0.00380
Within R^2 (Flat 350)	0.00871	0.01077	0.00784	0.00602

Note: Baseline model regression results for different race track lengths (each regression includes all three pollutants simultaneously). Although not reported here, each regression includes all control variables as specified in Equation 1. Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

Table 9. Pollutant-specific Regression Results for Flat 550, 330 and 350 with Fixed Effects

Dependent Variable:	Finish Time (s)			
Stage:	(1)	(2)	(3)	(4)
Flat 550				
PM_{10} (µg/m^3)	0.0017***	0.0019***	0.0016***	0.0018***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
O_3 (µg/m^3)	0.0008***	0.0005***	0.0005***	3.27×10^{-7}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
NO_2 (µg/m^3)	-0.0007*	0.0008*	0.0007*	0.0005
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Flat 330				
PM_{10} (µg/m^3)	0.0009**	0.0010***	0.0010***	0.0008**
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
O_3 (µg/m^3)	0.0004**	0.0004**	0.0005***	0.0004*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
NO_2 (µg/m^3)	0.0007	0.0008	0.0010	0.0008
	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Flat 350				
PM_{10} (µg/m^3)	0.0017***	0.0019***	0.0016***	0.0018***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
O_3 (µg/m^3)	0.0006***	0.0004***	0.0007***	0.0007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
NO_2 (µg/m^3)	-0.0019***	-0.0008*	-0.0010*	-0.0010**
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Fixed-effects				
Dog	Yes	Yes	Yes	Yes
Stadium		Yes	Yes	Yes
Season			Yes	Yes
Year				Yes
Fit statistics				
Observations (Flat 550)	32,455	32,455	32,455	32,455
Observations (Flat 330)	24,511	24,511	24,511	24,511
Observations (Flat 350)	27,329	27,329	27,329	27,329
R^2 (Flat 550; PM_{10})	0.68827	0.69503	0.69606	0.69845
R^2 (Flat 550; O_3)	0.68860	0.69513	0.69616	0.69843
R^2 (Flat 550; NO_2)	0.68827	0.69505	0.69608	0.69844
R^2 (Flat 330; PM_{10})	0.63338	0.63361	0.63409	0.63829
R^2 (Flat 330; O_3)	0.63336	0.63358	0.63413	0.63827
R^2 (Flat 330; NO_2)	0.63327	0.63348	0.63397	0.63822
R^2 (Flat 350; PM_{10})	0.63056	0.63515	0.63678	0.63930
R^2 (Flat 350; O_3)	0.63047	0.63485	0.63680	0.63921
R^2 (Flat 350; NO_2)	0.63048	0.63474	0.63652	0.63896
Within R^2 (Flat 550; PM_{10})	0.01531	0.02473	0.01932	0.01680
Within R^2 (Flat 550; O_3)	0.01635	0.02504	0.01965	0.01672
Within R^2 (Flat 550; NO_2)	0.01529	0.02478	0.01937	0.01677
Within R^2 (Flat 330; PM_{10})	0.00685	0.00717	0.00438	0.00354
Within R^2 (Flat 330; O_3)	0.00678	0.00710	0.00446	0.00346
Within R^2 (Flat 330; NO_2)	0.00653	0.00683	0.00405	0.00332
Within R^2 (Flat 350; PM_{10})	0.00717	0.01021	0.00687	0.00511
Within R^2 (Flat 350; O_3)	0.00692	0.00941	0.00690	0.00485
Within R^2 (Flat 350; NO_2)	0.00695	0.00911	0.00614	0.00419

Note: Pollutant-specific regression results for different race track lengths (for each race track length, the coefficient of each pollutant was estimated in a separate regression). Although not reported here, each regression includes all control variables as specified in Equation 1. Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: $p<0.01$, **: $p<0.05$, *: $p<0.1$*

A.4 Pollutant-Specific Regression without Outliers

Table 10. Pollutant-Specific Results without Outliers

Dependent Variable:	Finish Time (s)		
Independent Variable:	PM_{10}	O_3	NO_2
Variables			
PM_{10} (µg/m^3)	0.0004^{***}		
	(0.0001)		
O_3 (µg/m^3)		0.0005^{***}	
		(5.37×10^{-5})	
NO_2 (µg/m^3)			-9.79×10^{-5}
			(0.0001)
Weight (kg)	0.0150^{***}	0.0149^{***}	0.0149^{***}
	(0.0014)	(0.0014)	(0.0014)
Track Condition (Value)	0.2159^{***}	0.2170^{***}	0.2162^{***}
	(0.0097)	(0.0097)	(0.0097)
Temperature (°C)	-0.0051^{***}	-0.0055^{***}	-0.0050^{***}
	(0.0002)	(0.0002)	(0.0002)
Precipitation (mm)	0.0220^{***}	0.0209^{***}	0.0213^{***}
	(0.0017)	(0.0017)	(0.0017)
Relative Humidity (%)	-0.0004^{***}	-0.0002^{***}	-0.0004^{***}
	(7.81×10^{-5})	(7.91×10^{-5})	(7.8×10^{-5})
Wind Speed (kt)	0.0039^{***}	0.0035^{***}	0.0039^{***}
	(0.0002)	(0.0002)	(0.0002)
Predominant Wind Direction (°)	$-6.77 \times 10^{-5***}$	$-6.66 \times 10^{-5***}$	$-7.73 \times 10^{-5***}$
	(9.1×10^{-6})	(8.74×10^{-6})	(8.99×10^{-6})
Fixed-effects			
Dog	Yes	Yes	Yes
Stadium	Yes	Yes	Yes
Season	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	376,356	376,382	376,378
R^2	0.57294	0.57317	0.57297
Within R^2	0.00704	0.00731	0.00700

Note: This table shows the estimated coefficients for the regressions with only one pollutant at a time as independent variable. For each regression, outlier values for the specific pollutant have been excluded from the data set based on visual inspection. Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

A.5 Selected Regression Results including Odds Ratio Control Variable

Table 11. Full Model Results including Odds Ratio Control Variable

Dependent Variable:	Finish Time (s)			
Stage:	(1)	(2)	(3)	(4)
Variables				
PM_{10} (μg/m^3)	0.0004***	9.26×10^{-5}	0.0002	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
O_3 (μg/m^3)	0.0010***	0.0011***	0.0010***	0.0006***
	(5.82×10^{-5})	(5.74×10^{-5})	(6.18×10^{-5})	(6.22×10^{-5})
NO_2 (μg/m^3)	-0.0014***	0.0009***	0.0011***	0.0007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Weight (kg)	0.0127***	0.0129***	0.0116***	0.0146***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Odds Ratio (Value)	-0.0006	-0.0010***	-0.0009**	-0.0002
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Track Condition (Value)	-0.0199***	0.2921***	0.3058***	0.2185***
	(0.0066)	(0.0102)	(0.0101)	(0.0097)
Temperature (°C)	-0.0059***	-0.0059***	-0.0054***	-0.0054***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Precipitation (mm)	0.0176***	0.0207***	0.0202***	0.0212***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)
Relative Humidity (%)	-0.0002***	-0.0003***	-0.0002**	-0.0002**
	(7.9×10^{-5})	(7.8×10^{-5})	(7.99×10^{-5})	(7.96×10^{-5})
Wind Speed (kt)	0.0026***	0.0038***	0.0035***	0.0036***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Predominant Wind Direction (°)	$-5.48 \times 10^{-5***}$	$-4.07 \times 10^{-5***}$	$-5.07 \times 10^{-5***}$	$-5.44 \times 10^{-5***}$
	(9.47×10^{-6})	(9.43×10^{-6})	(9.41×10^{-6})	(9.36×10^{-6})
Fixed-effects				
Dog	Yes	Yes	Yes	Yes
Stadium		Yes	Yes	Yes
Season			Yes	Yes
Year				Yes
Fit statistics				
Observations	373,526	373,526	373,526	373,526
R^2	0.56255	0.56950	0.57111	0.57336
Within R^2	0.00996	0.01432	0.01098	0.00739

Note: Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

Table 12. Pollutant-Specific Results with Fixed Effects including Odds Ratio Control Variable

Dependent Variable:	Finish Time (s)		
Independent Variable:	PM_{10}	O_3	NO_2
Variables			
PM_{10} (µg/m^3)	0.0003***		
	(0.0001)		
O_3 (µg/m^3)		0.0005***	
		(5.37×10^{-5})	
NO_2 (µg/m^3)			4.58×10^{-6}
			(0.0001)
Weight (kg)	0.0147***	0.0146***	0.0147***
	(0.0014)	(0.0014)	(0.0014)
Odds Ratio (Value)	-0.0002	-0.0002	-0.0002
	(0.0004)	(0.0004)	(0.0004)
Track Condition (Value)	0.2159***	0.2169***	0.2160***
	(0.0097)	(0.0097)	(0.0097)
Temperature (°C)	-0.0051***	-0.0055***	-0.0051***
	(0.0002)	(0.0002)	(0.0002)
Precipitation (mm)	0.0223***	0.0213***	0.0216***
	(0.0017)	(0.0017)	(0.0017)
Relative Humidity (%)	-0.0004***	-0.0002***	-0.0004***
	(7.84×10^{-5})	(7.93×10^{-5})	(7.82×10^{-5})
Wind Speed (kt)	0.0039***	0.0034***	0.0039***
	(0.0002)	(0.0002)	(0.0002)
Predominant Wind Direction (°)	-7.04×10^{-5}***	-6.91×10^{-5}***	-7.83×10^{-5}***
	(9.13×10^{-6})	(8.77×10^{-6})	(9×10^{-6})
Fixed-effects			
Dog	Yes	Yes	Yes
Stadium	Yes	Yes	Yes
Season	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	373,526	373,526	373,526
R^2	0.57322	0.57333	0.57320
Within R^2	0.00707	0.00732	0.00704

Note: This table shows the estimated coefficients for the regressions with only one pollutant at a time as independent variable. Standard errors in parenthesis are clustered at the entity (dog) level.
*Significance levels: ***: p<0.01, **: p<0.05, *: p<0.1*

References

Aguilar-Gomez, S., Dwyer, H., Graff Zivin, J., and Neidell, M. (2022). This is air: The "nonhealth" effects of air pollution. *Annual Review of Resource Economics*, 14:403–425.

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3):91–93.

Anderson, J. O., Thundiyil, J. G., and Stolbach, A. (2012). Clearing the air: a review of the effects of particulate matter air pollution on human health. *Journal of medical toxicology*, 8:166–175.

Araneda, O. F. (2022). Horse racing as a model to study the relationship between air pollutants and physical performance. *Animals*, 12(9):1139.

Araneda, O. F. and Cavada, G. (2022). Atmospheric pollutants affect physical performance: A natural experiment in horse racing studied by principal component analysis. *Biology*, 11(5):687.

Austin, W., Heutel, G., and Kreisman, D. (2019). School bus emissions, student health and academic performance. *Economics of Education Review*, 70:109–126.

Beavan, A., Härtel, S., Spielmann, J., and Koehle, M. (2023). Air pollution and elite adolescent soccer players' performance and well-being; an observational study. *Environment International*, 175:107943.

Behm, D. G. and Carter, T. B. (2021). Empathetic factors and influences on physical performance: a topical review. *Frontiers in Psychology*, 12:686262.

Bigazzi, A. Y. and Figliozzi, M. A. (2014). Review of urban bicyclists' intake and uptake of traffic-related air pollution. *Transport Reviews*, 34(2):221–245.

Bodor, Z., Bodor, K., Keresztesi, Á., and Szép, R. (2020). Major air pollutants seasonal variation analysis and long-range transport of pm 10 in an urban environment with specific climate condition in transylvania (romania). *Environmental Science and Pollution Research*, 27:38181–38199.

CAMS (2022). Institut national de l'environnement industriel et des risques (ineris), aarhus university, norwegian meteorological institute (met norway), jülich institut für energie- und klimaforschung (iek), institute of environmental protection – national research institute (iep-nri), koninklijk nederlands meteorologisch instituut (knmi), meteo france, nederlandse organisatie voor toegepast-natuurwetenschappelijk onderzoek (tno), swedish meteorological and hydrological institute (smhi), finnish meteorological institute (fmi), italian national agency for new technologies, energy and sustainable economic development (enea) and barcelona supercomputing center (bsc): Cams european air quality forecasts, ensemble data. `https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-europe-air-quality-reanalyses?tab=overview`. [Accessed 06-02-2024].

Chen, H., Lin, Y., Su, Q., and Cheng, L. (2017). Spatial variation of multiple air pollutants and their potential contributions to all-cause, respiratory, and cardiovascular mortality across china in 2015–2016. *Atmospheric environment*, 168:23–35.

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: A bad mix. *Technical Note, Duke University*, 7.

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale university press.

Datcu, R. F., Brîndescu, S., and Petracovschi, S. (2021). Anxiety and athlete performance: A systematic narrative review of the mutual influence of these concepts. *Timisoara Physical Education and Rehabilitation Journal*, 14(26):62–75.

de Vocht, F., Katikireddi, S. V., McQuire, C., Tilling, K., Hickman, M., and Craig, P. (2021). Conceptualising natural and quasi experiments in public health. *BMC medical research methodology*, 21:1–8.

Doncheva, N. T., Palasca, O., Yarani, R., Litman, T., Anthon, C., Groenen, M. A., Stadler, P. F., Pociot, F., Jensen, L. J., and Gorodkin, J. (2021). Human pathways in animal models: possibilities and limitations. *Nucleic acids research*, 49(4):1859–1871.

Elminir, H. K. (2005). Dependence of urban air pollutants on meteorology. *Science of the total environment*, 350(1-3):225–237.

EPA (2017). Air quality in ireland 2016. indicators of air quality. `https://www.epa.ie/publications/monitoring--assessment/air/Air-Quality-In-Ireland-2016.pdf`. [Accessed 06-02-2024].

EPA (2023). Ireland's ambition to move towards the health-based who air quality guidelines will be challenging, but will have a significantly positive impact on health. `https://www.epa.ie/news-releases/news-releases-2023/irelands-ambition-to-move-towards-the-health-based-who-air-quality-guidelines-will-be-challenging-but-will-have-a-significantly-positive-impact-on-health.php`. [Accessed 06-02-2024].

EPA (2024). Ireland's environment: Air. `https://www.epa.ie/our-services/monitoring--assessment/assessment/irelands-environment/air/`. [Accessed 07-05-2024].

European Parliament and Council of the European Union (2008). Directive 2008/50/ec of the european parliament and of the council of 21 may 2008 on ambient air quality and cleaner air for europe.

Fu, S., Viard, V. B., and Zhang, P. (2021). Air pollution and manufacturing firm productivity: Nationwide estimates for china. *The Economic Journal*, 131(640):3241–3273.

Gabryś, H. S., Gote-Schniering, J., Brunner, M., Bogowicz, M., Blüthgen, C., Frauenfelder, T., Guckenberger, M., Maurer, B., and Tanadini-Lang, S. (2022). Transferability of radiomic signatures from experimental to human interstitial lung disease. *Frontiers in medicine*, 9:988927.

Gates, M. C. (2007). The influence of air pollution on thoroughbred race performance. *Equine and Comparative Exercise Physiology*, 4(2):79–88.

Gatto, N. M., Henderson, V. W., Hodis, H. N., John, J. A. S., Lurmann, F., Chen, J.-C., and Mack, W. J. (2014). Components of air pollution and cognitive function in middle-aged and older adults in los angeles. *Neurotoxicology*, 40:1–7.

Geukes, K., Harvey, J. T., Trezise, A., and Mesagno, C. (2017). Personality and performance in real-world competitions: Testing trait activation of fear of negative evaluation, dispositional reinvestment, and athletic identity in the field. *Psychology of Sport and Exercise*, 30:101–109.

GREY2K (2024). Greyhound racing around the world. `https://www.grey2kusa.org/about/worldwide.php`. [Accessed 01-02-2024].

GRI (2024a). About greyhound racing ireland. `https://www.grireland.ie/About-GRI/about-gri/`. [Accessed 05-02-2024].

GRI (2024b). Racing nights and times. `https://www.grireland.ie/racing/race-nights--times/`. [Accessed 06-02-2024].

Guo, M. and Fu, S. (2019). Running with a mask? the effect of air pollution on marathon runners' performance. *Journal of Sports Economics*, 20(7):903–928.

Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. CRC Press.

IFHA (2024). International federation of horseracing authorities - facts and figures. `https://www.ifhaonline.org/default.asp?section=Resources&area=4`. [Accessed 01-02-2024].

Kampa, M. and Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2):362–367.

Kayes, I., Shahriar, S. A., Hasan, K., Akhter, M., Kabir, M., and Salam, M. (2019). The relationships between meteorological parameters and air pollutants in an urban environment. *Global Journal of Environmental Science and Management*, 5(3):265–278.

Langley, G., Evans, T., Holgate, S. T., and Jones, A. (2007). Replacing animal experiments: choices, chances and challenges. *Bioessays*, 29(9):918–926.

Le Tertre, A., Medina, S., Samoli, E., Forsberg, B., Michelozzi, P., Boumghar, A., Vonk, J., Bellini, A., Atkinson, R., Ayres, J., et al. (2002). Short-term effects of particulate air pollution on cardiovascular diseases in eight european cities. *Journal of Epidemiology & Community Health*, 56(10):773–779.

Lichter, A., Pestel, N., and Sommer, E. (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics*, 48:54–66.

Liu, Y., Zhou, Y., and Lu, J. (2020). Exploring the relationship between air pollution and meteorological conditions in china under environmental governance. *Scientific reports*, 10(1):14518.

Marcora, S. M., Staiano, W., and Manning, V. (2009). Mental fatigue impairs physical performance in humans. *Journal of applied physiology*.

Matt, F., Cole-Hunter, T., Donaire-Gonzalez, D., Kubesch, N., Martínez, D., Carrasco-Turigas, G., and Nieuwenhuijsen, M. (2016). Acute respiratory response to traffic-related air pollution during physical activity performance. *Environment international*, 97:45–55.

MET (2024). Climate of ireland. https://www.met.ie/climate/climate-of-ireland. [Accessed 18-02-2024].

Mullins, J. T. (2018). Ambient air pollution and human performance: Contemporaneous and acclimatization effects of ozone exposure on athletic performance. *Health economics*, 27(8):1189–1200.

Oireachtas (2023). Houses of the oireachtas: Dáil Éireann debate - wednesday, 29 nov 2023. https://www.oireachtas.ie/en/debates/debate/dail/2023-11-29/16/. [Accessed 05-02-2024].

Oxford Stadium (2024). How do you calculate racing odds in greyhounds? https://oxford-stadium.co.uk/blog/how-do-you-calculate-racing-odds-in-greyhounds/. [Accessed 16-04-2024].

Pincus, S. and Stern, A. C. (1937). A study of air pollution in new york city. *American Journal of Public Health and the Nations Health*, 27(4):321–333.

Poole, D. C. and Erickson, H. H. (2011). Highly athletic terrestrial mammals: horses and dogs. *Comprehensive Physiology*, 1(1):1–37.

RCÉ (2023). Rásaíocht con Éireann: Annual report 2022. https://www.grireland.ie/Resource/reports-and-statistics/annual/. [Accessed 05-02-2024].

Sarkar, M. and Fletcher, D. (2013). How should we measure psychological resilience in sport performers? *Measurement in Physical Education and Exercise Science*, 17(4):264–280.

Shehab, M. and Pope, F. (2019). Effects of short-term exposure to particulate matter air pollution on cognitive performance. *Scientific reports*, 9(1):8237.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2):39–42.

Tainio, M., Andersen, Z. J., Nieuwenhuijsen, M. J., Hu, L., De Nazelle, A., An, R., Garcia, L. M., Goenka, S., Zapata-Diomedi, B., Bull, F., et al. (2021). Air pollution, physical activity and health: A mapping review of the evidence. *Environment international*, 147:105954.

Täubert, H., Agena, D., and Simianer, H. (2007). Genetic analysis of racing performance in irish greyhounds. *Journal of Animal Breeding and Genetics*, 124(3):117–123.

Towcester Racecourse (2024). What are the different race distances in greyhound racing? https://towcester-racecourse.co.uk/what-are-the-different-race-distances-in-greyhound-racing/. [Accessed 16-02-2024].

Tsujino, I., Kawakami, Y., and Kaneko, A. (2005). Comparative simulation of gas transport in airway models of rat, dog, and human. *Inhalation toxicology*, 17(9):475–485.

Van Cutsem, J., Marcora, S., De Pauw, K., Bailey, S., Meeusen, R., and Roelands, B. (2017). The effects of mental fatigue on physical performance: a systematic review. *Sports medicine*, 47(8):1569–1588.

Wendler, A. and Wehling, M. (2010). The translatability of animal models for clinical development: biomarkers and disease models. *Current opinion in pharmacology*, 10(5):601–606.

Zhang, Y., Ke, L., Fu, Y., Di, Q., and Ma, X. (2022). Physical activity attenuates negative effects of short-term exposure to ambient air pollution on cognitive function. *Environment International*, 160:107070.

Zivin, J. G. and Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–3673.

Zuo, C., Chen, J., Zhang, Y., Jiang, Y., Liu, M., Liu, H., Zhao, W., and Yan, X. (2023). Evaluation of four meteorological reanalysis datasets for satellite-based pm2. 5 retrieval over china. *Atmospheric Environment*, 305:119795.

Declaration of consent

on the basis of Article 30 of the RSL Phil.-nat. 18

Name/First Name: Weibel Peter

Registration Number: 17-111-675

Study program: MSc in Climate Sciences

Bachelor ☐ Master ☑ Dissertation ☐

Title of the thesis: Impact of air pollution on greyhound racing performance

Supervisor: Prof. Dr. Eric Strobl

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of 5 September, 1996 is authorized to revoke the title awarded on the basis of this thesis.

For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with future theses submitted by others.

Bern, 22.05.2024

Place/Date

Signature