Data-driven analysis of the impact of $PM_{2.5}$ exposure on respiratory flow in asthmatic adolescents

Filip Futera

4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh

2024

Abstract

Air pollution poses a serious threat to human health, especially for people with respiratory diseases like asthma. Inhaling these airborne pollutants can exacerbate asthma symptoms and induce changes in respiratory flow by causing conditions like dyspnea. Although existing studies have associated air pollution with changes in certain components of respiratory flow, they have not established definitive causal relations between them. This is because they were unable to control for confounding factors when analysing these relations and also because no metric currently exists to measure the regularity of respiratory flow. This study addresses this gap by investigating and quantifying the causal relations between $PM_{2.5}$ and respiratory flow in asthmatic adolescents while accounting for temperature, humidity, activity level and sleep. It uses respiratory, physical activity and environmental datasets from the DAPHNE project and, for the first time, develops a medically-informed metric for approximating the regularity of respiratory flow. The findings reveal, for the first time, significant causal links between $PM_{2.5}$ and respiratory flow in most of the asthmatics adolescents studied. They also show that the effect of $PM_{2.5}$ on respiratory flow is significant when contrasted with physical activity level, temperature, humidity and previous respiratory flow, and exhibits considerable variation in both direction and magnitude across patients, time lags and months.

Research Ethics Approval

The ethical approval for the DAPHNE study was granted by the Institute Ethics Committee of the All India Institute for Medical Science, Delhi (Reference numbers: IEC-256/05.05.2017, RP-26/2017, OP13/03.08.2018).

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Filip Futera*)

Acknowledgements

I would like to express my sincerest gratitude to my supervisor Professor D K Arvind for his mentorship and continued support throughout this project. His valuable insights, guidance, and advice have been instrumental in shaping this dissertation.

In addition, I would like to greatly appreciate Dr Gordon Drummond for his support and advice throughout this journey. His medical insights were key in directing my research methodology and interpreting results.

I would also like to thank Professor Michael Gutmann for his role as the second marker for my project and the feedback he has given me, which helped me refine my research methodology.

I would finally like to thank my family and friends for their continued support and motivation throughout not only the dissertation but also the entire university experience.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

A study conducted in 2017 highlighted the uptrend in respiratory disease morbidity, demonstrating a 39.8% increase since 1990 to 544.9 million affected individuals in 2017 [75]. The study also identified chronic respiratory diseases to be responsible for 7% of all global mortalities in 2017, marking an 18% increase since 1990 [75]. These findings raise a pivotal question: despite the advancements in technology and healthcare, why does the burden of respiratory conditions continue to grow? Although answering this question is not the objective of this investigation, researching these conditions and their underlying causes helps us better understand them and mitigate their burden on society.

Dysfunctional breathing, characterized by episodes of irregular respiratory flow, is observed in 9.5% of healthy adults and is more than twice as prevalent in asthma patients [83]. It has a significant comorbidity with high BMI and poor physical condition in children, and is linked to breathlessness that can result in functional impairment and increased anxiety [3, 42]. In asthma patients, irregular breathing patterns during sleep have been linked to increased nocturnal asthmatic symptoms [11]. This highlights the relationship between irregular respiratory flow and asthma severity. Yet, despite its significant impact on health, abnormal breathing continues to be poorly understood and has no gold standard for its diagnosis [83, 3]. It is also often misdiagnosed by healthcare professionals, leading to frequent sub-optimal treatment [83]. Investigating the causes of irregular respiratory flow can therefore help doctors diagnose and manage abnormal breathing, leading to more targeted and effective treatment of respiratory diseases.

Meanwhile, rising air pollution levels [70] have increasingly shown a growing link to symptoms of respiratory distress [57, 5], such as coughing [74]. Investigating the relationship between $PM_{2.5}$ and irregular respiratory flow can therefore help medical professionals better understand these respiratory conditions. It can also help predict periods of abnormal breathing, which would not only facilitate preventive treatment but also aid in reducing the incidence and severity of respiratory diseases.

1.2 Project Aims and Contributions

This study aims to analyze datasets on air pollution, respiratory and physical activities collected in the DAPHNE project by the Centre for Speckled Computing at the University of Edinburgh. The goal of this is to investigate the health effects of air pollution on respiratory functions. This study in particular investigates the causal relationship between $PM_{2.5}$ exposure and the regularity of respiratory flow in asthmatic adolescents using the PCMCI+ causal discovery algorithm [66]. It then quantifies this causal exposure-response relationship using causal effect estimation [68].

The study introduces the following contributions for the first time:

- Developed a medically-informed metric that measures the regularity of respiratory flow in a non-invasive manner.
- Examined the causal relationship between $PM_{2.5}$ and respiratory flow in asthmatic patients across different time lags over 1-hour and 8-hour periods after exposure.
- Accounted for the impact of sleeping on the cause-response relationship of $PM_{2.5}$ and respiratory flow.
- Quantified the direct and indirect causal effects of $PM_{2.5}$ on the regularity of respiratory flow.

To make the above contributions, multiple conceptual problems were solved throughout the investigation. Firstly, with no gold standard method existing to measure respiratory flow or even the abnormality of breathing episodes [3], developing a metric for approximating the regularity of respiratory flow was particularly challenging. It required an integration of domain-specific knowledge to ensure the metric's accuracy and frequent refinements in consultation with a medical professional. Additionally, accurately identifying breaths, approximating their tidal volume, and extracting their features to be used by this metric posed significant conceptual difficulties. Finally, incorporating sleep as a variable in the causal discovery stage was challenging. Controlling for sleep is critical to isolating the effects of $PM_{2.5}$ exposure on respiration and ensuring accurate results, however, its absence in related studies made it difficult to model.

1.3 Report Outline

This report covers the analysis outlined in Section 1.2 and is structured as follows. **Chapter 2** provides the relevant background knowledge and examines related research, using their limitations to clarify the research goals. **Chapter 3** explores the methodology used to accurately identify breaths, extract their features, and integrate the resultant respiratory dataset with the environmental data. **Chapter 4** outlines and justifies the metric developed to measure the regularity of respiratory flow. **Chapter 5** establishes and examines the causal relations between $PM_{2.5}$ and the regularity of respiratory flow. **Chapter 6** then estimates and analyses the effect strengths of the causal relations established in Chapter 5. Finally, the main conclusions are critically analyzed in **Chapter 7** alongside a discussion of limitations and future work.

Chapter 2

Background

2.1 DAPHNE Project

The Delhi Air Pollution: Health and Effects (DAPHNE) project is a collaboration between nine institutions across the UK and India [25]. The project collected respiratory, physical activity and environmental datasets from asthmatic adolescents, aged 10 to 18, in Delhi India across 2018-2020 [25]. The data was collected using two non-invasive miniature wearable sensors that were worn by the patients: the RESpeck and AIRSpeck sensors. Both sensors were developed by the Centre for Speckled Computing at the University of Edinburgh.

The RESpeck is a 45x38x13 mm device that contains tri-axial accelerometer and gyroscope sensors [18]. The sensor is worn by each patient on the lower rib cage using a piece of tape to keep the sensor position and orientation fixed. The tri-axial accelerometer measurements are continually transmitted to the patient's local device for storage using Bluetooth LE technology [18]. These readings are then automatically processed to compute the values outlined in Table 3.1 that are then used in this investigation. The RESpeck's most important for this study measurement is the breathing signal (see Figure 3.1 for an example). It is a physiological signal representing the respiratory activity of the patient and is analogous to a flow rate waveform [9]. It is the core component used in this investigation to analyse respiratory flow. The signal is computed by aggregating the smoothed angle changes of the tri-axial measurements relative to gravity [9]. This makes use of the fact that during respiration, the chest wall on which the RESpeck is located expands and changes shape. This in turn causes the angle of the device to change relative to gravity [9], thereby tracking the respiratory activity.

The AIRSpeck is a personal clip-on device that is worn on clothing [25]. It is part of the AirSpeck family of personal and static sensors [6] that measure airborne pollution particles. The AIRSpeck uses an Optical Particle Counter sensor to collect data on the concentrations of particulate matter (PM) in the air [6]. Particulate matter is a collection of different chemical pollutants. It consists of microscopic solids or liquid droplets like dust, dirt and smoke that become hazardous when inhaled [24]. PM is usually grouped by cumulative particle diameter size. A popular grouping is $PM_{2.5}$ that refers to the fine inhalable particles with diameters of 2.5 micrometers and smaller [24]. The AIRSpeck

measures PM concentrations for 16 bins corresponding to separate particle diameters in the range of 0.38µm to 17µm [6]. It uses this to then compute the aggregated $PM_{2.5}$ concentrations that are used in this study. The AIRSpeck device also uses a Sensirion SHT21 sensor to measure the ambient temperature and humidity [6].

2.2 Related Work

2.2.1 Air Pollution and Respiration

The consequences of air pollution on general human health have been widely researched in recent years. Air pollution caused by human activity accounts for the death of approximately 9 million people yearly [58]. This is a key motivating factor for research into better understanding and mitigating the effects that air pollution may have on health. Although air pollution has been most prominently associated with respiratory conditions, it was also shown to impact other organs and systems. A study showed that air pollution can cause neuronal cell damage that leads to permanent brain damage and neurological diseases like Alzheimers [29, 47].

However, the most common consequence of air pollution is its short-term and long-term impact on respiratory functions [30]. Several studies have investigated the association between air pollution and respiratory conditions in the general population. Tran et al. linked increased air pollution with respiratory inflammation, oxidative stress and impairment of the lung's immune system functionality [81]. They emphasized that the association is particularly prevalent in vulnerable populations like children and older adults [81]. They warned that the inevitable climate change resulting from economic growth is bound to amplify air pollution levels [81] and urged for further research into the health effects of air pollution. Meanwhile, Glick et al. associated outdoor $PM_{2.5}$ and O_3 levels with pediatric pneumonia hospitalization and mortality in children across 2007-2008 [33].

In 2019 a study by Xie et al. directly compared the economic and health effects of $PM_{2.5}$ and ozone pollution in China. They discovered that $PM_{2.5}$ is associated with substantially more severe health effects, especially respiratory. However, they also found that $PM_{2.5}$ has a much more profound economic impact than ozone pollution. They predicted that without control policies, $PM_{2.5}$ could lead to a 2% loss in gross domestic product in China by 2030 [89]. They thereby argued that research and economic policies should focus on better understanding and mitigating the effects of $PM_{2.5}$ in particular [89].

Increasingly, research has been focusing on analysing the impact of air pollution on respiration in people with respiratory diseases. Several studies have demonstrated that people with respiratory diseases like asthma and chronic obstructive pulmonary disease (COPD) are more vulnerable to the effects of air pollution than the general population [89, 52]. There is therefore a more profound need to research and understand the effects of air pollution on health in these groups of people. Jiang et al. found that air pollution is associated with onset of asthma and increased respiratory disease morbidity and related mortality [44]. They also showed that the effect of air pollution on health in asthma

patients varies greatly by season, time and country [44]. Tiotiu et al. found that exposure to outdoor air pollutants, including $PM_{2.5}$, can lead to increased asthma symptoms and also decreased lung function in asthmatic children [80]. Meanwhile, Ng et al. associated a $2.64\mu g/m^3$ increase in traffic-sourced $PM_{2.5}$ with a 0.33% decrease in peak expiratory flow in severe asthmatic patients in Nagasaki, Japan [61]. The peak expiratory flow measures the maximal flow rate that can be achieved during expiration. It is therefore a component of respiratory flow and a method of checking lung function [7]. It is also often indicative of impending asthma attacks or underlying respiratory conditions [7]. The study therefore directly correlated increased $PM_{2.5}$ exposure with decreased lung function in asthmatic adults. These findings were further supported in a study by Edginton et al. that used random effect models to associate a $10\mu g/m^3$ increase in $PM_{2.5}$ with a $1.02L/min$ decrease in peak expiratory flow among non-smoking asthmatics [20].

Clearly, a lot of research has already associated air pollution with adverse respiratory outcomes, particularly decreased lung function, in both asthmatic and non-asthmatic people. However, these studies have limitations. Firstly, they predominantly focus on lung function, often approximated using peak expiratory flow. This is a limitation because peak expiratory flow only measures the maximal speed of expiration. Therefore, these lung function measures fail to account for various key components of the respiratory cycle like how quickly and deeply an individual breathes. Moreover, Eid et al. showed that peak expiratory flow is often an inaccurate measure of respiratory function in asthmatic patients, where air trapping may cause the peak expiratory flow to give misleading reassurance of normal function [21]. Secondly, these studies between air pollution and respiratory function only discover correlatory associations between the two. They do not establish causal relationships between air pollution and respiratory outcomes let alone quantify them. Moreover, they do not account for confounding factors when associating air pollution with adverse health effects. Several confounding factors have been shown to exacerbate respiratory function (see Section 2.2.2). Accounting for these factors while investigating the associations between air pollution and respiratory outcomes would offer more accurate insights into the impact of air pollution on respiration. This makes the existing studies limited in the insights they offer.

That being said, a small number of studies have established significant causal relations between $PM_{2.5}$ and respiratory outcomes in asthmatics [57, 5, 74] while accounting for confounding factors. However, they too are limited. They focus on respiratory rate and coughing outcomes rather than general respiratory dynamics like respiratory flow. There is therefore a clear research gap on the effects of air pollution on more complete measures of respiratory dynamics in asthmatics, particularly through causal inference.

This study addresses this gap by focusing on the relation between air pollution and respiratory flow, an area previously unexplored in existing literature. Respiratory flow refers to the dynamics behind respiration, encompassing the rate, volume, and speed of air movement in and out of the lungs during breathing. It is therefore a more complete measure of respiratory function and dynamics than just peak expiratory flow, which is, in fact, also a component of respiratory flow itself. This study also further fills the research gap by focusing on the causal relations between air pollution and respiratory flow in asthmatics, while accounting for key confounding factors. It thereby addresses

the identified lack of causal inference and control for confounding factors in related work, by employing a robust causal analysis. Nonetheless, the related work motivates the specifics of this study. It has identified asthmatics as substantially more vulnerable to the adverse health effects of air pollution than healthy people [52]. It also demonstrated $PM_{2.5}$ to have more severe health and economic impacts than other pollutants [89]. Subsequently, this study focuses on investigating the causal relations between $PM_{2.5}$ and respiratory flow in asthmatics, while controlling for several confounding factors.

2.2.2 Confounding Factors and Respiration

Related work has found increasingly more confounding factors that influence the respiratory activities of asthmatics. To accurately identify the causal relations between $PM_{2.5}$ and respiratory flow in this study, it is important to account for these confounding factors to ensure the results are valid. In this subsection, related studies that identify these confounding factors are overviewed and discussed in the context of this investigation.

Environmental factors have been demonstrated to impact respiration in asthmatic patients. Li et al. found that ambient temperature was negatively correlated with both peak expiratory flow rate and forced expiratory volume in asthmatic children [53]. They used this to associate an increase in ambient temperature with a decrease in lung function among asthmatic children [53]. Similarly, Chen et al. showed a U-shaped trend between ambient temperature and asthma related hospitalizations [13]. They demonstrated that both very warm and very cold temperatures generally increase asthma symptom severity. Several other studies have identified causal relations between temperature and changes in respiratory rate or coughing [57, 5, 74]. Hayes et al. found that breathing hot humid air triggered bronchoconstriction (narrowing of lung airways often resulting in shortness of breath) in asthmatic patients [39]. They also found that asthmatic patients have a lower forced expiratory volume when respiring humid air at room temperature [39]. Humidity is also generally correlated with ambient temperature and weather conditions, and has been shown to affect asthma symptoms in patients [56]. In summary, studies have demonstrated that temperature and humidity impact respiration and lung function [39, 74, 5, 53]. Humidity and temperature are therefore accounted for in this study to ensure the effects of $PM_{2.5}$ on respiratory flow are accurately isolated and appropriately analysed.

There are also other environmental factors that have been shown to impact respiration. Osborne et al. associated pollen exposure with asthma symptom exacerbations leading to hospitalizations [62]. Liu et al. found that various allergens, including dust and pollen, impact respiration [54]. Exposure to these allergens typically decreases lung function (measured as peak expiratory flow or forced expiratory volume) in asthmatics [54]. That being said, these allergens were not included in the study. This is because their highly variable nature makes them difficult to measure accurately and they were also not collected as part of the DAPHNE project.

There are also many non-environmental factors that impact respiration. Physical activities, particularly those of high intensity, directly impact respiration in both asthmatic and non-asthmatic people [73]. However, even smaller activity changes like changing lying or sitting position can impact the respiratory flow of asthmatics. Admirabilis et al.

showed that changing position to the supine sleep position reduces asthma symptoms during the night [2]. Interestingly, sleep itself has also been demonstrated to have an impact on asthma symptoms and respiration. A study by Bohadana et al. described the tendency of asthma to get worse at night, often due to nocturnal increase in airway inflammation and bronchial responsiveness [1]. Therefore, both physical activity intensity (level) and sleeping are accounted for in this study.

In summary, there are several environmental and non-environmental factors that have been associated with changes in respiration and lung function. These include temperature, humidity and activity level, which are accounted for in this study to ensure the relation between $PM_{2.5}$ and respiratory flow is accurately evaluated. Sleep is another factor that should be accounted for in these studies to ensure valid results. However, existing research typically overlooks accounting for sleep when investigating the causal relation between $PM_{2.5}$ and respiratory outcomes [57, 5, 74]. This study overcomes this limitation by running a parallel causal analysis on the impact of $PM_{2.5}$ on respiratory flow with an additional sleep variable.

2.2.3 Respiratory Flow

There is a big gap in literature where no work has attempted to measure respiratory flow let alone approximate its regularity. Doing so can offer more comprehensive insights into respiratory function and quality of breathing, that directly relate to asthma symptoms and underlying conditions. Related work has however attempted to quantify certain components of respiratory flow like tidal volume.

To measure tidal volume, most studies use expert equipment like a spirometer or plethysmograph that is not easily accessible and is not user-friendly [38, 34, 65]. Nitrogen washout and gas dilution have also been used for this task [85], but again also require expert equipment that is not user-friendly. Other studies place strict requirements on data collection to ensure their method of computing tidal volume is applicable. A study by Wang et al. used nocturnal respiratory sounds including snoring to measure tidal volume. However, this approach only worked during sleep. A limited number of studies have attempted less restrictive approaches that utilize more accessible equipment. Lujan et al. used the area under the flow waveform curve to approximate the tidal volume [55] while Miller et al. used the forced expiratory volume [59]. Although these techniques have shown to be effective and user-friendly approximations of tidal volume, they still fall short of accurately capturing the complex dynamics of respiratory flow that extend beyond just tidal volume.

It is therefore evident that there is still a gap on measuring respiratory flow dynamics in literature. This study addresses this gap by developing the first medically-informed metric that approximates the regularity of respiratory flow, leveraging data from miniature wearable sensors.

Chapter 3

Data Pre-Processing

3.1 Overview

The data used in this investigation comprises of 220 AIRSpeck and RESpeck recordings from 137 asthmatic adolescents. The RESpeck was used to continuously compute values overviewed in Table 3.1 using the established methods in [9] as outlined in Section 2.1. Meanwhile, the AIRSpeck device continuously collected the ambient temperature, humidity and $PM_{2.5}$ readings (among other fields) in each patient's direct vicinity.

Field	**Description**
Timestamp	Exact time and date of recording
Breathing Signal	Physiological signal analogous to the flow rate waveform
Respiratory Rate	Number of breaths taken per minute
Activity Level	Intensity of the patient's activity
Activity Type	The category of the activity the patient is carrying out

Table 3.1: Data collected by RESpeck device adapted from [5]

Approximately 37% of the RESpeck data is missing [5]. This accounts for episodes when the patient is non-stationary (walking, running and other movements) or the device is not worn. These episodes are picked up by the activity level being above 0.3 and below 0.013 respectively [5]. The data for non-stationary activities is excluded because the breathing signal becomes inaccurate and unreliable during high-intensity activities.

3.2 Breath Capturing

To accurately analyze respiratory flow patterns, it is crucial to first process the raw RESpeck data and compute features for each breath. These features are then combined with AIRSpeck data in Section 3.4 for a comprehensive causal analysis in Chapter 5. The key challenge in this process is designing an algorithm capable of accurately detecting individual breaths from the breathing signal in the RESpeck data. This task

is complicated because of sensor noise and occasional signal offsets that are present in the recorded breathing signals. Initially a naive approach was used to treat any zero-crossing in the breathing signal as a transition between exhalation and inhalation. This approach considered a breath to begin at the first zero-crossing and end at the third. However, breath visualisations exemplified in Appendix A.1 showed that this method was susceptible to misinterpreting sensor noise and signal offsets as spurious breaths. It resulted in frequent occurrences of marginal zero-crossings being misidentified as several breaths and prolonged periods of signal offsets being misinterpreted as a single long breath.

Subsequently, the ”gold standard” method established and implemented in [26] was adapted to this task to overcome these issues. This method was tested in [26] against a nasal cannula to ensure its reliability and accuracy in capturing breaths. It employs a sliding window technique on the breathing signal to provide a constantly updated positive and negative threshold based on the root mean square of the signal's waveform amplitude within the window [26]. The breathing signal must cross this threshold to be recognized as either a new exhalation or inhalation. The threshold is also bounded by a minimum value that ensures it never reaches zero. To further optimize this approach, experiments were run with different hyper-parameters. For each experiment random samples of breaths for each patient were visualised to ensure they are not obscured by sensor noise and signal offsets. Ultimately, a 30s window size and 0.001 minimum threshold were selected as they captured the least noisy breaths among those sampled. A sample breathing signal decomposition into breaths with this method is shown in Figure 3.1. Meanwhile, Appendix A.2 shows breaths captured from Patient DAP134's breathing signal, where this ”gold-standard” method correctly handles sensor noise.

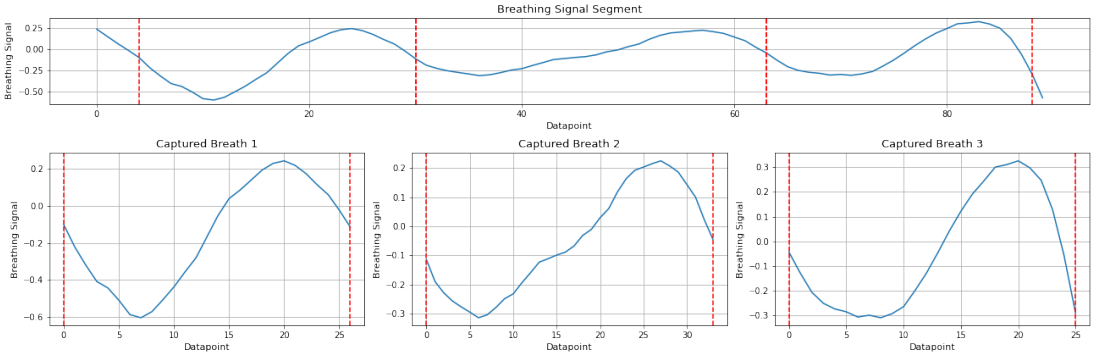


Figure 3.1: Example Decomposition of Breathing Signal Into Breaths For Patient DAP102(1)

3.3 Breath Feature Extraction

Quantifying respiratory flow is a complex task. Literature often uses respiratory rates when investigating the impact of a cause on breathing [57, 5]. Respiratory rate is also useful in identifying respiratory symptoms like tachypnea (abnormally high respiratory rate) [63]. Therefore, the first breath feature was the respiratory rate averaged across the 30-second window around each breath. This averaging process helped mitigate potential

errors and inaccuracies that sometimes occur in the RESpeck's outputs. However, respiratory rate alone is not sufficient to represent respiratory flow. It overly focuses on the breath frequency and offers little information about the quality of the breath itself. To address this, further features were computed to approximate the tidal volume of a breath. This included the area under the breathing signal curve and peak respiratory flow. The area under the curve was computed using the trapezoidal rule for numerical integration shown in Formula 3.1. Alternatives to this method like the Simpson rule were considered. However, the breathing signal curve is composed of a finite number of data points, connected by linear segments. This structure made the trapezoidal rule a more directly suitable choice for integration, due to its inherent suitability for linear segments [82]. Meanwhile, the peak respiratory flow was measured as the maximum absolute value of the breathing signal in the breath.

$$\text{Area} \approx \sum_{i=2}^{n} \frac{1}{2}(f_i + f_{i-1})(t_i - t_{i-1}) \tag{3.1}$$

where:

- n is the number of samples captured in the breath,
- $f = (f_1, f_2, \ldots, f_n)$ is the breathing signal,
- $t = (t_1, t_2, \ldots, t_n)$ is the timestamp.

Tidal volume is a key determinant of healthy respiration [38] and may be used to identify irregular, labored breathing [32]. Therefore approximating tidal volume is essential for accurately quantifying the regularity of respiratory flow. There is a strong correlation between peak respiratory flow, area under the breathing signal curve and true tidal volume [87, 23]. In fact, area under the breathing signal curve has been used to approximate tidal volume in related tasks of extracting respiratory parameters from speech recordings [60]. It is therefore an apt approximation of tidal volume to use in this investigation. These are the reasons why each breath's area, peak respiratory flow and localized mean respiratory rate were used to approximate the regularity of respiratory flow in Chapter 4. Other features were experimented with, including counting the number of local minima and maxima in each breath's signal curve. The hypothesis was that more irregular breaths will have more local extrema. However, this approach was too sensitive to sensor noise and was therefore excluded from the derived feature-set.

3.4 Data Fusion

Having processed the RESpeck recordings to identify breaths and compute their corresponding features, the next step was to fuse the AIRSpeck and RESpeck data together. This was done on a per-breath basis using Algorithm 1, where the breath data-point was merged with the closest AIRSpeck data-point, at a cut-off of 30 seconds. The mean time delay between consecutive AIRSpeck entries across patients is less than 52 seconds in the dataset. Therefore a 30-second buffer in both directions in Algorithm 1 is adequate to locate the closest AIRSpeck point, if available. However, beyond this

threshold, the accuracy of temperature, humidity, and $PM_{2.5}$ values for a given breath cannot be ensured. The alternative of using the closest *n* AIRSpeck points for a given breath was attempted. However, due to the high variability of $PM_{2.5}$ concentrations that can occur within just a few minutes, this was a less accurate approach. Table 3.2 outlines the final fields in the fused AIRSpeck and processed RESpeck datasets.

Algorithm 1 Fusing AIRSpeck and RESpeck Data

Input: RESpeck breaths dataset, AIRSpeck environmental data
Output: Fused dataset
fusedDataset ← new Dataset()
for each breath in RESpeck dataset **do**
    timestamp ← (breath.startTimestamp + breath.endTimestamp) / 2
    closestEntry ← closest AIRSpeck data point to timestamp by time
    if |(closestEntry.timestamp - timestamp).total_seconds()| ≤ 30 seconds **then**
        fusedEntry ← breath + closestEntry
    else
        fusedEntry ← breath + NaNs
    end if
    fusedDataset.add(fusedEntry)
end for

Field	Description
patient	Patient ID and recording number
startTimestamp	Exact time and date of breath begin
endTimestamp	Exact time and date of breath end
timestamp	Exact time and date of breath midpoint
peakRespiratoryFlow	Maximum absolute breathing signal in breath
area	Area under breathing signal curve
meanBreathingRate	Mean respiratory rate in 30sec window around breath
meanActivityLevel	Mean intensity of patient's activity during breath
modeActivityType	Most frequent category of activity the patient is performing
temperature	AIRSpeck ambient temperature
humidity	AIRSpeck ambient humidity
PM2_5	AIRSpeck ambient $PM_{2.5}$ concentration
Sleeping Features	Features to estimate probability of sleeping (section 5.6)

Table 3.2: Fused Data from RESpeck and AIRSpeck

3.5 Outlier Removal

As is common when working with sensor data, sensor noise and sensor malfunction create outliers that must be removed. However, in this case detecting outliers is tricky. When dealing with tidal volume features it is important to distinguish between outliers and highly irregular breaths. The line between the two is often blurry. Therefore as little breath related outliers were filtered as possible. By visualising the results of Section

3.4, Algorithm 2 was used to handle outliers on a per-patient basis. It removes breaths with duration beyond 5 seconds or with feature values that are negative or outside the upper threshold established using the 1.5xIQR rule.

Algorithm 2 RESpeck Breath Filtering Algorithm

Input: Dataset of breath features
Let $Q3(var)$ be the 3rd quartile value of variable *var*.
Let $IQR(var)$ be the inter-quartile range of variable *var*.
Let PRF be the peak respiratory flow of a breath.
Calculate $Q3$ and IQR for peak respiratory flow and area
for each breath in dataset **do**
    if breath.duration $>$ 5 seconds **then**
        Remove breath from dataset
    end if
    if breath.area ≤ 0 **or** breath.PRF ≤ 0 **then**
        Remove breath from dataset
    end if
    if breath.PRF $> Q3(\text{PRF}) + 1.5 \times IQR(\text{PRF})$ **or** breath.area $> Q3(\text{area}) + 1.5 \times IQR(\text{area})$ **then**
        Remove breath from dataset
    end if
end for

Typical inhalations last 1-1.5 seconds and exhalations 1.5-2 seconds [40]. Therefore it is safe to exclude breaths longer than 5 seconds. Although the 1.5xIQR outlier removal rule is used to remove upper outliers, breaths with features less than $Q1 - 1.5IQR$ (first quartile minus 1.5 times the inter-quartile range) were not excluded. This is because such breaths are possible, often arising in episodes of dyspnea (condition of breathlessness and "air hunger" [83]), which is prevalent among asthmatic patients [91]. Meanwhile, outliers from AIRSpeck recordings that exceeded the lower and upper cutoff thresholds in Table 3.3 were replaced by NaNs.

Finally, two patient recordings were entirely excluded from the study. Patient DAP029's only recording contained $PM_{2.5}$ readings in the range 0 to 0.087469, with 65.44% of the data having a $PM_{2.5}$ concentration of 0. Similarly, the first recording for Patient DAP002 contains temperature entries that are all recorded as 0. With both of these variables being essential components of the causal analysis, the study could not be conducted on those recordings. Subsequently the analysis in Chapter 5 was run on 136 patients with 218 total recordings.

Field	**Lower Bound**	**Upper Bound**
Temperature	1.0	50.0
Humidity	1.0	100.0
$PM_{2.5}$	0.1	1500.0

Table 3.3: Thresholds for outlier removal on AIRSpeck data

3.6 Missing Data Imputation

Sensor malfunction and outlier detection resulted in a lot of missing data in temperature, humidity and $PM_{2.5}$ readings. On average, 12.33% of a patient's RESpeck data lacks corresponding AIRSpeck data due to periods when the AIRSpeck was off while the RESpeck was recording alone. However, these sensor inaccuracies and outliers can be imputed accurately [5]. Previous research with the same dataset has conducted this imputation using linear interpolation with a window size of 5 in both directions [5]. This approach is limited by its assumption of a linear trend in $PM_{2.5}$ concentration through time. Recent studies have demonstrated the non-linear relationship of $PM_{2.5}$ with other correlated variables [79, 14]; therefore, this assumption cannot be made.

Instead, Multiple Imputation with Chained Equations (MICE) is used [8]. This approach takes the mean imputed value across multiple iterations, ensuring a more accurate result [8]. To avoid making assumptions about the trend of the imputed fields, Scikit-learn's RandomForestRegressor [64] is used. This approach allows for flexibility in capturing both linear and non-linear relationships among $PM_{2.5}$, temperature, and humidity over time. The imputation of temperature is done using the previous 5 and next 5 temperature entries, current humidity and hour of the day variables. Similarly, the humidity is imputed using the previous 5 and next 5 humidity entries, current temperature and hour of the day variables. Meanwhile, $PM_{2.5}$ is imputed using the previous 5 and next 5 $PM_{2.5}$ entries and the hour of the day variable. With weather and air pollution changing throughout the day as sun intensity varies [76], hour of the day is an important variable for imputing these fields. Importantly, if the imputation window for a variable consists entirely of NaNs, that variable remains unimputed to preserve accuracy. This ensures that the imputation process does not introduce artificial values based on insufficient data, especially during extended periods where the AIRSpeck was not recording.

To ensure the accuracy of the imputation, experiments were run with different hyper-parameters to select the MICE configuration with the best performance. Each experiment was run on the same sample of 20 randomly selected patients. For each patient, a random sample of 500 data points with non-missing entries for $PM_{2.5}$, temperature and humidity was selected. Their values were temporarily set to NaNs and imputed using the relevant MICE setup. The mean absolute percentage error (MAPE) of the imputation was then reported as an average across the selected patients. Table 3.4 shows the results of the best performing hyper-parameter combinations. It justifies selecting imputation with 10 iterations of 4 estimator random forests with max depth of 10.

Iterations	**Estimators**	**Max Depth**	**Temperature MAPE (%)**	**Humidity MAPE (%)**	**$PM_{2.5}$ MAPE (%)**
10	4	10	0.01	0.06	2.13
20	4	10	0.02	0.07	2.13
10	4	8	0.02	0.09	2.74

Table 3.4: MICE Imputation Experiment Results

Importantly, breath features were not imputed. It is difficult to do so reliably and it may artificially create episodes of irregular respiration that cause inaccurate results.

Chapter 4

Approximating Respiratory Flow

To examine the impact of $PM_{2.5}$ on respiratory flow, a metric must be established to quantify the regularity of respiratory flow within any given breathing episode. Measuring respiratory flow alone is not enough because every breath has a distinct flow, which does not necessarily imply a change in respiratory dynamics. Instead the regularity of these flows must be assessed. However, with no literature having done this before, this is conceptually a difficult task. To address this, Dr Gordon Drummond, a medical professional, was consulted throughout the metrics development to ensure its accuracy.

The starting point for the metric was to focus on approximating rapid-shallow breathing (see Appendix B.1 to understand pictorially) as more irregular. Rapid-shallow breathing can occur in episodes of dyspnea and tachypnea [63]. Both these conditions are recognized as dysfunctional and abnormal breathing [83]. Therefore it makes sense to characterise irregular respiratory flow as episodes of rapid-shallow breathing.

With this in mind, the respiratory features from Section 3.3 were used to quantify the regularity of each breath on a per-patient basis. Quantifying the regularity at breath-level, rather than analyzing aggregated respiratory flow across multiple breaths, provided more fine-grained insights into the variability of the respiratory cycles. It also allowed better fine-tuning of the regularity metric to match the rapid-shallow specification. The study still investigates the impact of $PM_{2.5}$ on respiratory flow across respiratory periods by resampling the computed regularity metric to 1 minute and 15 minute resolutions in Chapter 5.

To identify rapid-shallow breaths, the area and peak respiratory flow features were used. In consultation with Dr Gordon Drummond, it was determined that rapid breaths exhibit a large peak respiratory flow [7] and shallow breaths have a smaller area [51]. These features were therefore used in unsupervised learning to distinguish the regularity of breaths. Initial approximations were established with the following three methodologies:

- K-Means clustering: 3 and 5 clusters were tried, for each setup the clusters were ranked on their rapid-shallow nature by considering the mean area and peak respiratory flow of the breaths in each cluster [49].
- Anomaly detection with isolation forests [90]: breaths on average in more shallow

branches of decision trees are considered more irregular.

- Anomaly detection by distance to the feature distribution: breaths with a larger Mahalanobis distance (see Appendix B.4) are considered more irregular in accordance with [46].

These methodologies proved highly effective in identifying breaths with smaller area and larger peak respiratory flow as more irregular. However, they often identified breaths with small peak respiratory flow and small area as equally irregular. Generally, they struggled to distinguish rapid-shallow breaths from just shallow breaths. Therefore in further consultation with Dr Gordon Drummond, the mean respiratory rate described in Section 3.3 was added to the feature set and a hand-crafted domain-informed metric was developed (instead of generic unsupervised learning techniques). By introducing the frequency domain, the final metric described in Algorithm 3 was able to isolate rapid breaths and match the provided specifications.

Algorithm 3 Final Algorithm to Calculate Regularity of Respiratory Flow

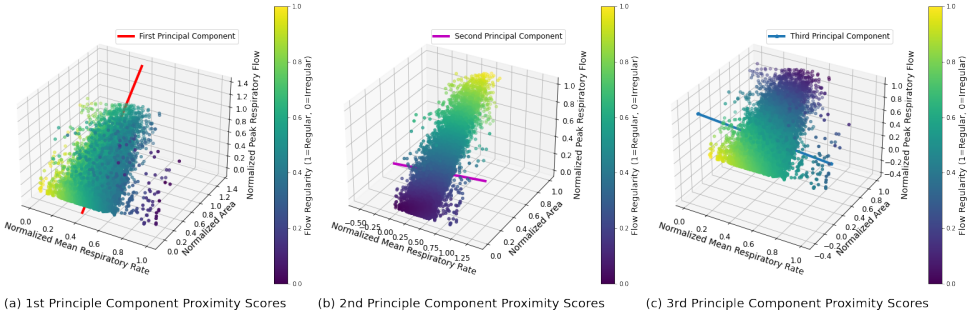
Input: Dataset of breath features df
Output: Metric score for the regularity of respiratory flow for each breath in df

procedure GETREGULARITY(df)
  Let $PC_i(vars)$ be the i'th principal component of PCA ran on variables $vars$
  Let PRF be the abbreviation for the peak respiratory flow variable set
  $df \leftarrow$ Normalize(df) ▷ Normalize into range [0,1]
  $d_0 \leftarrow \|df - (df \cdot PC_1(area, PRF)) \otimes PC_1(area, PRF)\|_2$
  $d_1 \leftarrow \|df - (df \cdot PC_1(area, PRF, BR\_mean)) \otimes PC_1(area, PRF, BR\_mean)\|_2$
  $d_2 \leftarrow \|df - (df \cdot PC_2(area, PRF, BR\_mean)) \otimes PC_2(area, PRF, BR\_mean)\|_2$
  $d_3 \leftarrow \|df - (df \cdot PC_3(area, PRF, BR\_mean)) \otimes PC_3(area, PRF, BR\_mean)\|_2$
  $d_{\text{total}} \leftarrow d_2 - d_0 - d_1 - d_3$ ▷ Consider each d_i an element of this metric
  return Normalize(d_{total}) ▷ Normalize into range [0,1]
end procedure

Algorithm 3 describes the final methodology used to approximate the regularity of each breath's respiratory flow. It is a combination of four elements, each calculated as the L2-norm of the difference between the data and its projection onto the respective principal components, derived from subsets of breath features. Principal Component Analysis (PCA) from Scikit-Learn [64] was used to do this. PCA reveals the underlying structure of the breath feature space, making its components key for differentiating regular from irregular breaths. To justify the four elements used in the metric, Figure 4.1 and Figure 4.2 show the effects of including each element for Patient DAP083(1). In these visualizations, data points marked with a lower proximity score[1], represented by a darker color shade, are indicative of more irregular breathing patterns. Conversely, points with a higher proximity score, denoted by a lighter color, are associated with more regular breathing patterns. Taking this into account, Figure 4.1 and Figure 4.2 show that:

[1]The proximity score is essentially the distance measured with L2-norm. However, since Algorithm 3 uses $1 - distance$ for some elements, proximity score is a more appropriate term.

- The proximity score to the first principal component of the distribution of area, peak respiratory flow and respiratory rate in Figure 4.1a effectively characterises breaths with high respiratory rate as irregular.
- The proximity score to the second principal component of the distribution of area, peak respiratory flow and respiratory rate in Figure 4.1b effectively characterises breaths with small area as irregular.
- The proximity score to the third principal component of the distribution of area, peak respiratory flow and respiratory rate in Figure 4.1c effectively characterises breaths with large peak respiratory flow as irregular.
- Although the proximity scores in Figure 4.1a associate high respiratory rate with irregular breaths, a lot of these irregular breaths also have small peak respiratory flow. Similarly, although the proximity scores in Figure 4.1c associate high peak respiratory flow with irregular breaths, a lot of these irregular breaths also have large area. This is also reflected in Figure 4.2b where some of the more irregular breaths have large respiratory rates and small area but also small peak respiratory flow. To address this, the proximity score to the first principal component of the distribution of area and peak respiratory flow in Figure 4.2a is used. It effectively characterises breaths with small area and high peak respiratory flow as irregular. Adding this element helps the algorithm converge on characterising truly rapid-shallow and high frequency breaths as irregular.

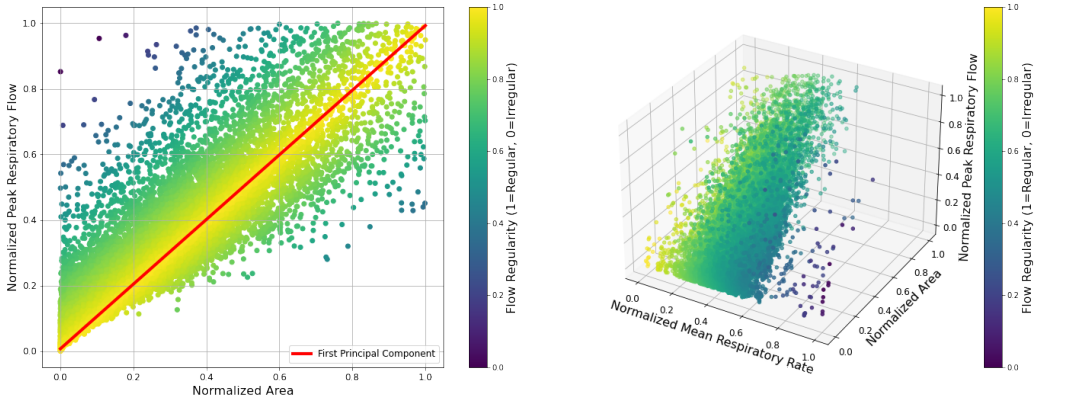


(a) 1st Principle Component Proximity Scores (b) 2nd Principle Component Proximity Scores (c) 3rd Principle Component Proximity Scores

Figure 4.1: Individual Metric Component Contributions for patient DAP083(1)

With these arguments in mind, each element in Algorithm 3 was meticulously chosen to align with the clinical definitions and, when aggregated, categorise breaths that are rapid-shallow and frequently occurring as irregular. Figure 4.3 confirms this approach. It visualises the distribution of area, peak respiratory flow and respiratory rate for patient DAP083(1) across different bins of regularity ranging from 0 (irregular) to 1 (regular). Breaths with lower values of respiratory flow regularity are observed to have smaller area, larger peak respiratory flow and higher breathing rate. This is also the case for other patients with further plots in Appendix B.3. Once again, this matches the medical expectations shared by Dr Gordon Drummond.

To justify generalising this metric to other patients, it is important to note that the patients have very similar distributions of breath features (see Appendix B.2). Therefore, the



(a) Proximity Score to 1st PC of Area and Peak Respiratory Flow

(b) Aggregated Proximity Scores to PCs of Area, Peak Respiratory Flow and Respiratory Rate

Figure 4.2: Further and Aggregated Metric Components for patient DAP083(1)

results in Figure 4.1 and Figure 4.2 were consistent across patients[2] and algorithm 3 yielded consistent and expected approximations for all patients.

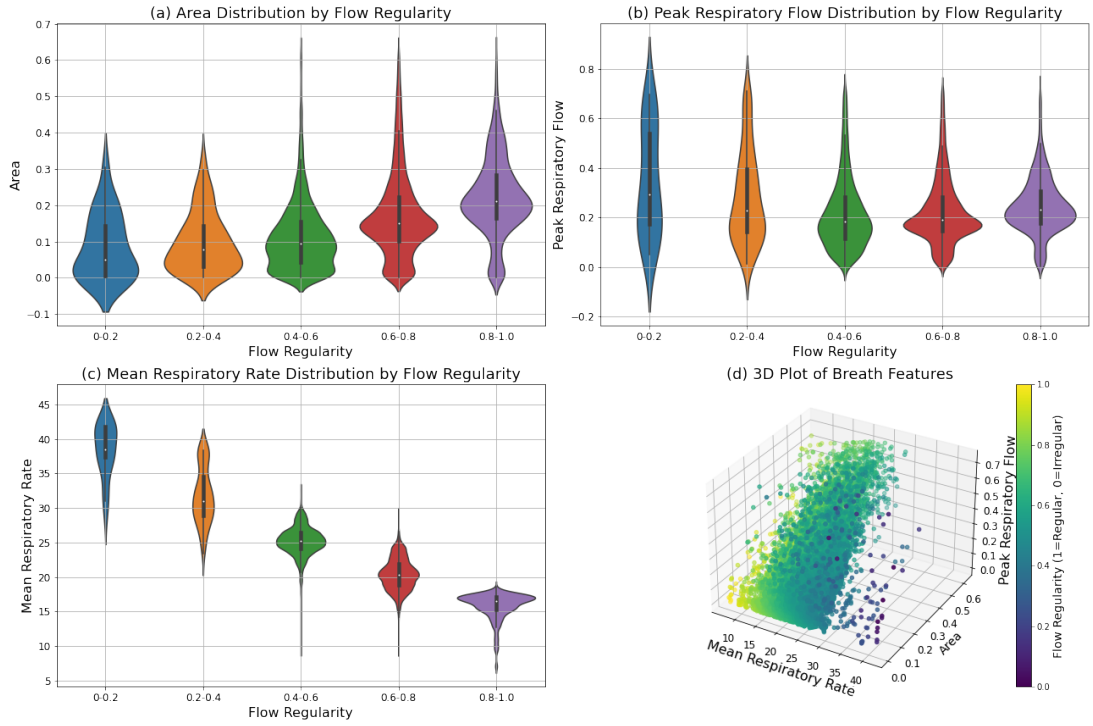


Figure 4.3: Distribution of breath features across different regularity bins for patient DAP083(1)

Finally, to better understand this metric, patterns between respiratory flow regularity and activity types are overviewed in Appendix B.5. Generally, there were no consistent trends between respiratory flow regularity and activity type across patients. This is because respiratory patterns across activity types varied substantially between patients.

[2]This was inferred when analysing the distributions of breath features before and after applying PCA for the patients. To showcase this, further patients' breath distributions and their principal components are showcased in Appendix B.2

Chapter 5

Causal Analysis

5.1 Regression Basis

Having processed the data and quantified the regularity of every breath's respiratory flow, the relationship between $PM_{2.5}$ and respiratory flow is investigated. However to warrant a causal investigation, the relationship between these two factors must first be established. This is done using Linear regression from Scikit-Learn [64], with and without $PM_{2.5}$, to predict the regularity of respiratory periods' flows. While polynomial regression and ensemble techniques like decision tree regressors were also attempted, they proved generally a much worse fit for this task.

Consequently, two regression models were established for each patient. The models were fitted on minute-averaged data for comparability to Section 5.3 and because of interest in irregular breathing episodes rather than individual breaths themselves. Each model aimed to predict the regularity of respiratory flow at time t using minute-average temperature, humidity and activity level up to time lag $t-60$ minutes. Additionally, the model with $PM_{2.5}$ had an extra set of minute-averaged $PM_{2.5}$ entries up to time lag $t-60$ minutes as well. The regressions were performed on a per-patient basis, with the mean adjusted R^2 values displayed in Figure 5.1. Adjusted R^2 was chosen over R^2 as it compensates for the addition of the $PM_{2.5}$ independent variables, preventing artificial goodness-of-fit inflation.

Figure 5.1 shows that adding $PM_{2.5}$ variables to the Linear regression model improves the mean adjusted R^2 score from 0.65 to 0.78. This implies that $PM_{2.5}$ is a significant variable that contributes to the explanation of the variance in the regularity of respiratory flow in patients. Therefore a relationship between $PM_{2.5}$ and the regularity of respiratory flow can be inferred. However, as with regressions, this relationship indicates correlation rather than causation. To establish a causal relationship between these variables, further analysis is needed using causal algorithms. Causal analysis, like in Chapter 5 and 6, is not only able to identify causal links, but also quantify them and determine the time lag at which these links occur. This is something that cannot be established from regression alone, therefore warranting a separate causal investigation.

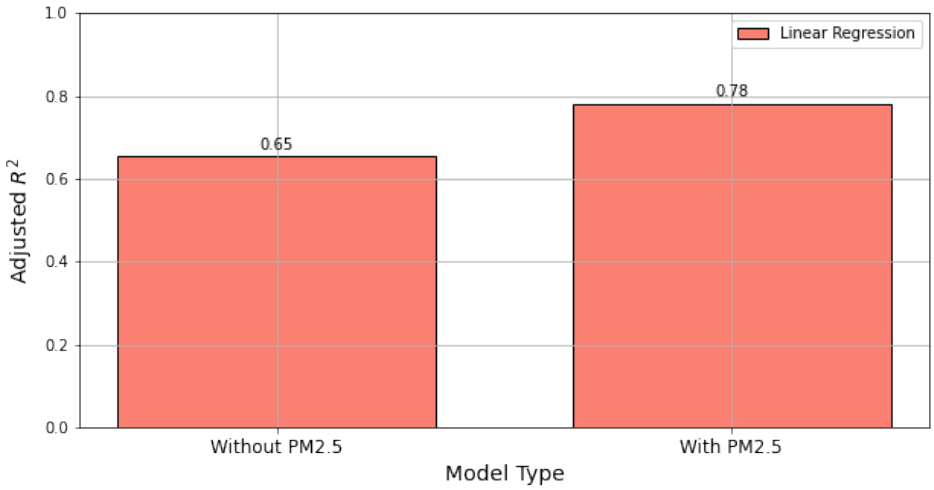


Figure 5.1: Adjusted R^2 of regressing respiratory flow regularity with and without $PM_{2.5}$ lags

5.2 Causal Algorithm

Having characterised the regularity of respiratory flow in Chapter 4, its causal relation with $PM_{2.5}$ was investigated next. To perform this analysis, a causal discovery algorithm must be used. The PCMCI+ algorithm [66] was used to do this. PCMCI+ discovers causal links in time-series data by testing for conditional independence between the variables in lagged (past to future) and contemporaneous (same-time) connections. It progressively removes links between variables that are conditionally independent given a conditioning set that it iteratively optimises. It is explained in more detail in Algorithm 4. PCMCI+ improves on the standard PCMCI algorithm by optimising the selection of conditioning sets in the conditional independence tests [66]. It is also more robust to the effects of auto-correlation, ensuring the detection of true causal links. While alternative causal discovery methods exist including PC [50], they often struggle with auto-correlation and thus suffer from inflated false positives [66]. Consequently, PCMCI+ is the most accurate algorithm for discovering contemporaneous (zero time lag) and lagged (time lag greater than zero) causal relations in time-series data.

To account for potential confounding factors and ensure a thorough analysis of the causal relationship between $PM_{2.5}$ and respiratory flow regularity, PCMCI+ was run with additional time-series of temperature, humidity and mean activity level for each patient. The algorithm was run twice on each patient, once for each study, using the Tigramite package [67]. The first study looked at the causal relations in the dataset at 1-minute resolution across a one-hour period ($\tau_{max} = 60$). Meanwhile, the second study used 15-minute resolution across an eight-hour period ($\tau_{max} = 32$). To achieve these resolutions, the dataset was resampled, taking the mean of each variable for the resolution period. Two studies were conducted to enable the analysis of both short-term and longer-term effects of exposure to $PM_{2.5}$ on the regularity of respiratory flow. This ensured a more complete study of the impact of $PM_{2.5}$ on respiratory flow.

Algorithm 4 High-Level PCMCI+ Algorithm with ParCorrWLS Conditional Test

Input: Time series dataset $\mathbf{X} = (X^1, \ldots, X^N)$, maximum time lag $\tau_{\max}$, significance threshold α_{PC}, conditional independence test ParCorrWLS$(\cdot,\cdot|\cdot)$.

function PCMCI+($\mathbf{X}, \tau_{\max}, \alpha_{\text{PC}}$)
  Let $\mathcal{P}(\mathcal{A})$ be a power set of an arbitrary set $\mathcal{A}$
  Let $\mathcal{G}$ be the time series graph with lagged and contemporaneous variables
  Let $B_t^-(X_t^j)$ be lagged parent links of X_t^j for $0 \leq \tau \leq \tau_{\max}$
  Initialize $\mathcal{G}$ with all lagged links from $B_t^-(X_t^j)$ for all $X_t^j \in \mathbf{X}$
  Initialize contemporaneous links in $\mathcal{G}$: Add $X_t^j \circ\!-\!\circ X_t^i$ for all $X_t^j \neq X_t^i \in \mathbf{X}$
  Let $\tilde{A}(X_t^j) := \{X_t^i \neq X_t^j \in \mathbf{X}\}$ be the contemporaneous adjacencies of X_t^j
  Let $p = 0$.
  for each adjacent pair $\{(X_{t-\tau}^i, X_t^j) \in \mathcal{G} : |\tilde{A}(X_t^j) \backslash \{X_{t-\tau}^i\}| \geq p, 0 \leq \tau \leq \tau_{\max}\}$ **do**
    for each $S \in \{Y \in \mathcal{P}(\tilde{A}(X_t^j) \backslash \{X_{t-\tau}^i\}) : |Y| = p\}$ **do**
      Set $Z := (S, B_t^-(X_t^j) \backslash \{X_{t-\tau}^i\}, B_{t-\tau}^-(X_{t-\tau}^i))$.
      Let (p-value, I) = ParCorrWLS$(X_{t-\tau}^i, X_t^j | Z)$
      if p-value $> \alpha_{\text{PC}}$ **then** ▷ Can't reject conditional independence given Z
        Delete link $X_{t-\tau}^i \to X_t^j$ from $\mathcal{G}$. ▷ Therefore no direct causal link
        break and continue to next adjacent pair
      end if
    end for
    Let $p = p + 1$.
    Re-compute $\tilde{A}(X_t^j)$ from $\mathcal{G}$ ▷ Then often sorted by minimum I value
  end for
  return $\mathcal{G}$.
end function

For each run of PCMCI+ a consistent set of variables was used. The significance threshold was set to $\alpha_{\text{PC}} = 0.02$. This is a common p-value threshold that is also used in related work in [57, 74]. Therefore, using this p-value also enabled the comparison of our results to results from related work with other respiratory outcomes in Section 5.5. The conditional independence (CI) test used was Partial Correlation Weighted Least Squares (ParCorrWLS). ParCorrWLS effectively deals with non-constant error variances by re-weighting data points based on their error variance [37]. It thereby offers better control over false positives, making it more accurate than other CI tests including ParCorr [37]. More specifically, it is designed to handle heteroskedasticity in time-series data, making it the suitable CI test choice for this analysis [37].

With this in mind, the PCMCI+ algorithm was then initialized and setup for this task (see Appendix C.1). It successfully ran on all 136 patients in the one hour study and 104 patients in the eight hour study, failing on patients where the data was too sparse.

5.3 Identifying Causal Links

Figure 5.2 illustrates the statistical significance of causal links from $PM_{2.5}$ to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution. At a p-value threshold of 0.02, 52.59% of the asthmatic patients exhibit a change in the regularity of their respiratory flow caused by exposure to $PM_{2.5}$. Among the patients whose respiratory flow is causally impacted by $PM_{2.5}$, there is an average of 1.549 significant causal links per patient. This value drops to an average of 0.815 causal links per patient from $PM_{2.5}$ to respiratory flow, when considering all the asthmatic patients. Meanwhile, at a still statistically significant p-value threshold of 0.05, 88.8% of patients have at least one causal link from $PM_{2.5}$ to respiratory flow, with a patient exhibiting on average 2.55 causal responses to $PM_{2.5}$ in their respiratory flow. These results show that $PM_{2.5}$ has a significant causal impact on respiratory flow over 1-minute periods, already within the first hour after exposure, in a majority of asthmatic patients.

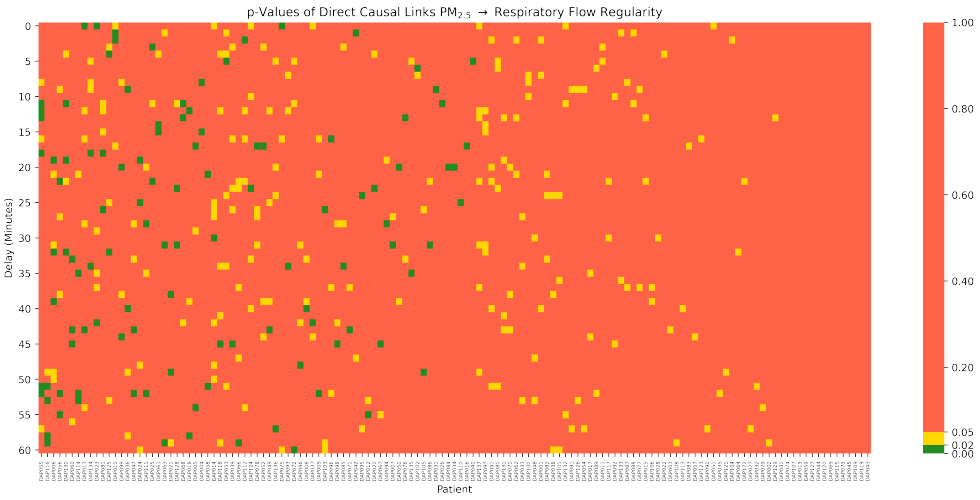


Figure 5.2: Causal Links of $PM_{2.5} \rightarrow$ Respiratory Flow Over 1-Hour Period

Having shown the prevalence of causal links from $PM_{2.5}$ to respiratory flow already within the first hour after exposure in Figure 5.2, Figure 5.3 extends this analysis to a longer time period. It shows the statistical significance of causal links from $PM_{2.5}$ to respiratory flow over 15-minute periods within the first 8 hours after exposure. At a p-value threshold of 0.02, 31.07% of patients have causal links from $PM_{2.5}$ to respiratory flow. Patients that exhibit causal links, are on average causally impacted at 1.469 time lags. With an increased p-value significant threshold of 0.05, 69.90% of patients have at least one causal link, with an average of 1.874 causal links per patient. The results therefore show that $PM_{2.5}$ still has a significant causal impact on respiratory flow over 15-minute periods, even up to eight hours after exposure, for many patients.

Comparing these results, there are significantly more patients with causal links between $PM_{2.5}$ and respiratory flow in the first hour after exposure (at 1-minute resolution), than at 15-minute resolution up to eight hours after exposure. This was also observed in the results of $PM_{2.5}$ links to respiratory rate and coughing in literature [69], and is likely attributed to two factors. Firstly, the impact of $PM_{2.5}$ likely diminishes over time as

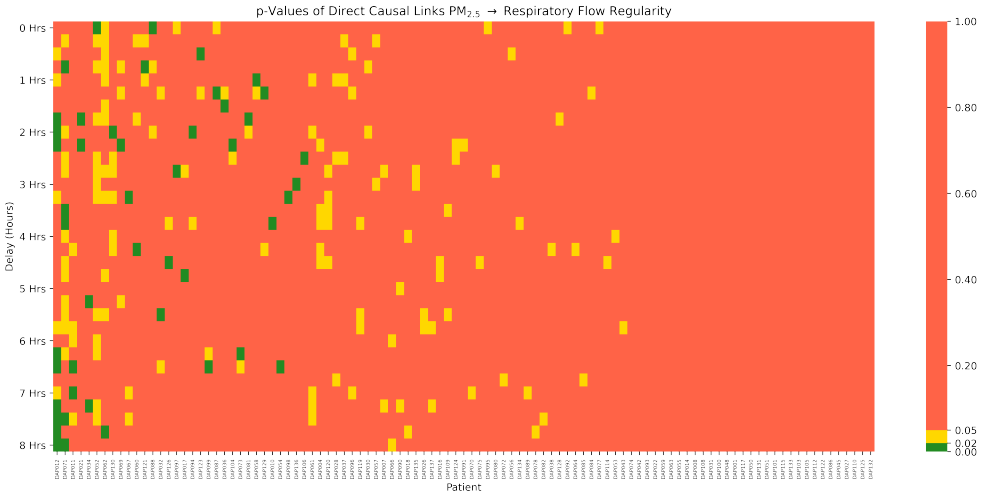


Figure 5.3: Causal Links of $PM_{2.5} \rightarrow$ Respiratory Flow Over 8-Hour Period

the body's natural defense mechanism mitigates the effect of $PM_{2.5}$ inhalation. In fact, the respiratory system's immediate physiological response to the inflammation [12] resulting from $PM_{2.5}$ inhalation [84] is likely more pronounced within the first hour after exposure. Secondly, respiratory flow's causal response to $PM_{2.5}$ exposure might be more subtle and continuously varying than expected. Unlike 1-minute respiratory flow periods, 15-minute periods may not capture these fine-grained, dynamic changes effectively. These factors could explain this disparity in causal link results between the 1-hour and 8-hour lag periods. However, further research would be needed to confirm these explanations. This might include investigating the changes of $PM_{2.5}$ concentration in the body with time and the subsequent reactions of the immune system.

5.4 Temporal Causal Link Analysis

The distribution of the number of causal links (NCL) between $PM_{2.5}$ and respiratory flow across time lags for both the 1-hour and 8-hour studies is displayed in Figure 5.4. It shows that the number of causal links is distributed quite uniformly with occasional peaks. In the 8-hour study, time lags 1.75 hours to 2.25 hours form a small cluster where more causal links are observed relative to the other time lags around them. This suggests that these lags are a period where the responses to $PM_{2.5}$ exposure are more consistently observed across patients.

Overall though, there are no clear dominating time lags at which significantly more patients have causal links. In the 1-hour study, the maximum number of causal links is 6 in time lag 52-minutes. In the 8-hour study it is 4 links at time lags 2.25 hours and 6.5 hours. Interpreting the results, this relatively uniform spread of causal links across the time lags makes sense. Everybody's immune system and physiological response mechanism is different [12]. In fact, this quite uniform distribution of links across lags was also observed in causal links between $PM_{2.5}$ and respiratory rate and coughing in [74, 5, 57, 69]. Therefore patients having causal links at different time lags matches expectations and highlights this disparity in every person's respiratory response to

$PM_{2.5}$ exposure.

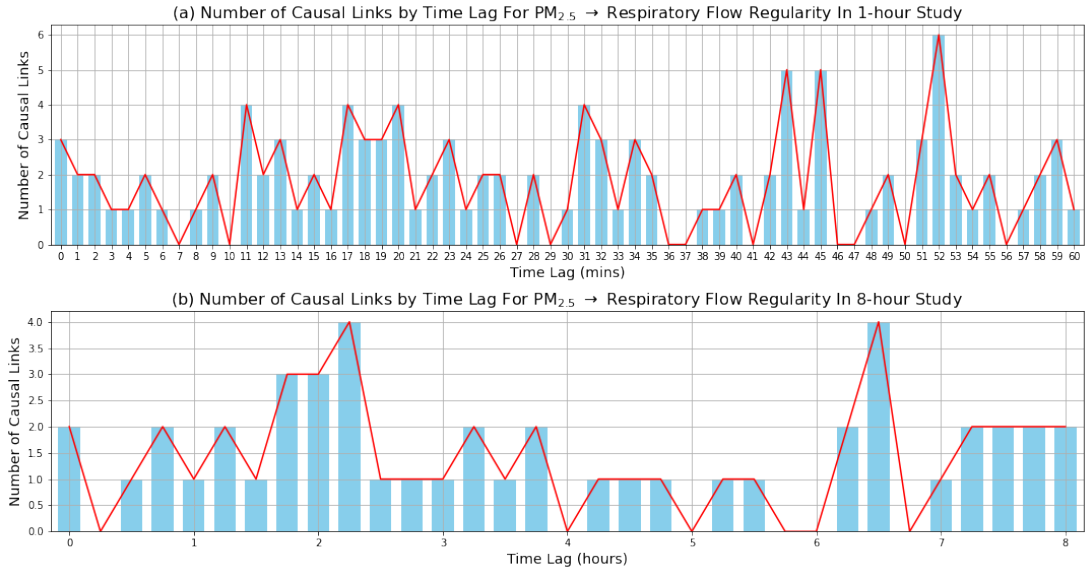


Figure 5.4: Number of $PM_{2.5} \rightarrow$ Respiratory Flow Causal Links Across Time Lags

Although Figure 5.4 is able to capture general temporal patterns in causal links, it does not offer insights into the trends of duration until patients exhibit their first causal link. To do this, survival analysis is performed using the Kaplan-Meier estimate [48]. It investigates the time lags at which a patient first exhibits a causal link. Figure 5.5 shows these results as the probability of a patient not exhibiting a causal link from $PM_{2.5}$ to respiratory flow across time lags for both the 1-hour and 8-hour studies separately.

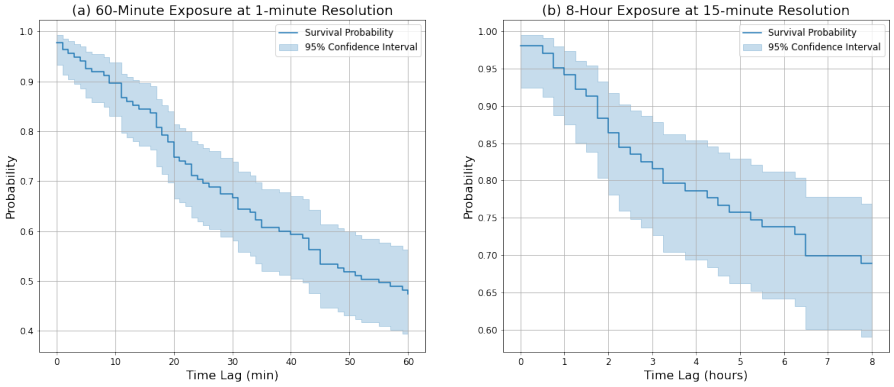


Figure 5.5: Survival Analysis: Probability of Patients Not Being Causally Impacted by $PM_{2.5}$ at 1-hour vs. 8-hour Max Lags With 95% Confidence Interval

Both studies show a steadily decreasing probability estimate curve for the majority of time lags. That being said, in the latter time lags, both curves show a diminishing rate of probability decline. The minute-level incidence rate of a patient exhibiting their first causal link after 45-60 minutes is much smaller than in the first 30 minutes after exposure. This is despite there still being a significant amount of patients with no causal

links yet. To be precise, this diminished rate of probability decline is most visible in lags 35-60 minutes. This means that although the number of causal links is relatively uniformly distributed in Figure 5.4, the causal links in the later time lags often belong to patients that already exhibited causal links previously. As such, it may be concluded that a patient is more likely to experience their first respiratory flow reaction to $PM_{2.5}$ within the first 35 minutes of exposure than afterwards. Finally, a similar trend can be seen in the 8-hour study. Patients are again more likely to have their first respiratory flow reaction, at 15-minute resolution, to $PM_{2.5}$ in the 0 to 2 hour lags than in the 6 to 8 hour lags.

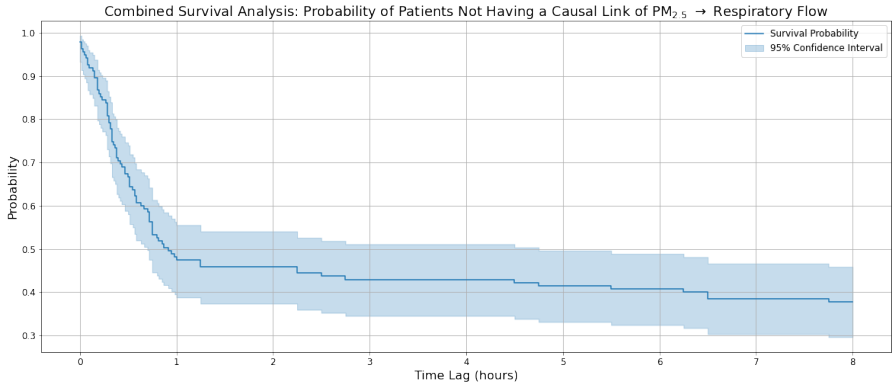


Figure 5.6: Survival Analysis: Probability of Patients Not Being Causally Impacted by $PM_{2.5}$ With Combined 1-hour and 8-hour Studies With 95% Confidence Interval

For a more complete visualisation, the individual survival analysis results in Figure 5.5 are combined into a single probability estimate curve. The 1-hour study results populate the first 60 minutes of the probability estimate curve with the remaining 7 hours coming from the 8-hour study. This combined curve is visualised in Figure 5.6. Firstly, it shows that 62.96% of all patients exhibit at least one causal link in either the 1-hour or 8-hour studies. It also further supports the observations from Figure 5.5. The incidence rate of patients exhibiting their first respiratory flow reaction to $PM_{2.5}$ exposure again clearly diminishes with time.

Figure 5.6 also shows several prolonged periods after the first hour, during which no patients exhibit their first causal reaction. Such periods are not visible in Figure 5.5. This means that a lot of patients that have causal links in the period between 1 hour to 7 hours after exposure also exhibit more fine-grained causal links in the first 60-minutes after exposure. In fact, 50% of patients with causal links in the 8-hour study, also have causal links in the 1-hour study.

To conclude this section, Figure 5.7 depicts the number of patients with $PM_{2.5}$ to respiratory flow causal links in each month, normalized by the number of patients with data in that month. Importantly, no data was collected in May as the patients were on school holiday. Otherwise, at least four patients had data collected in each of the other months. In the 1-hour study, the majority of patients with data in December and July have causal links. Meanwhile, in March, June and October, the 8-hour study has the highest ratio of patients with causal links to the number of patients with data in those

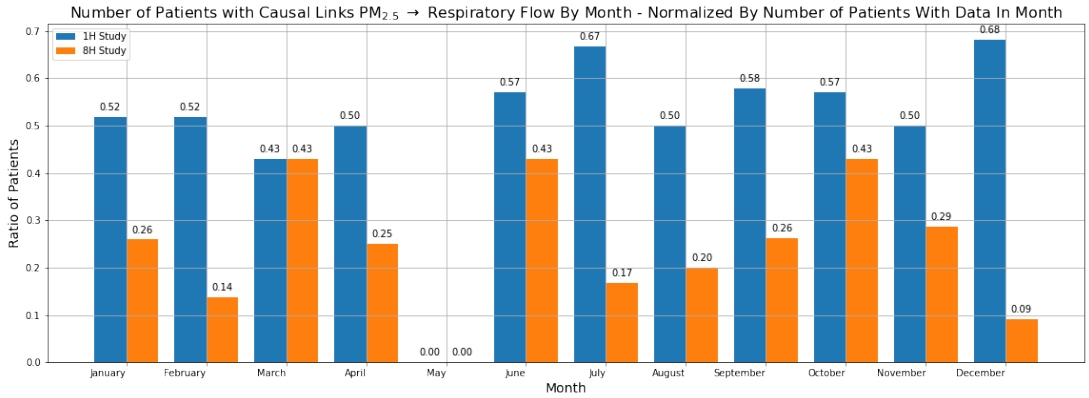


Figure 5.7: Number of Patients $PM_{2.5} \rightarrow$ Causal Links Across Months

months at 0.43. This could suggest that patients are more sensitive to $PM_{2.5}$ in those months and hence more frequently affected. However, it could also just be an artefact of patient heterogeneity, where the patients in those months were simply more susceptible to having a respiratory reaction to $PM_{2.5}$. Interestingly, out of all the months, December has the highest ratio of patients with causal links in the 1-hour study but the lowest ratio of patients with links in the 8-hour study. This shows that in December the asthma patients have much quicker and more fine-grained respiratory flow responses to $PM_{2.5}$ exposure, where just 9% of them have longer-term links from the 8-hour study while 68% of them have short-term links from the 1-hour study.

5.5 Comparison of Respiratory Outcome Results

Having established the causal relations between $PM_{2.5}$ and respiratory flow in Section 5.3, the results are compared to related DAPHNE studies on the causal impact of $PM_{2.5}$ on different respiratory outcomes in [69].

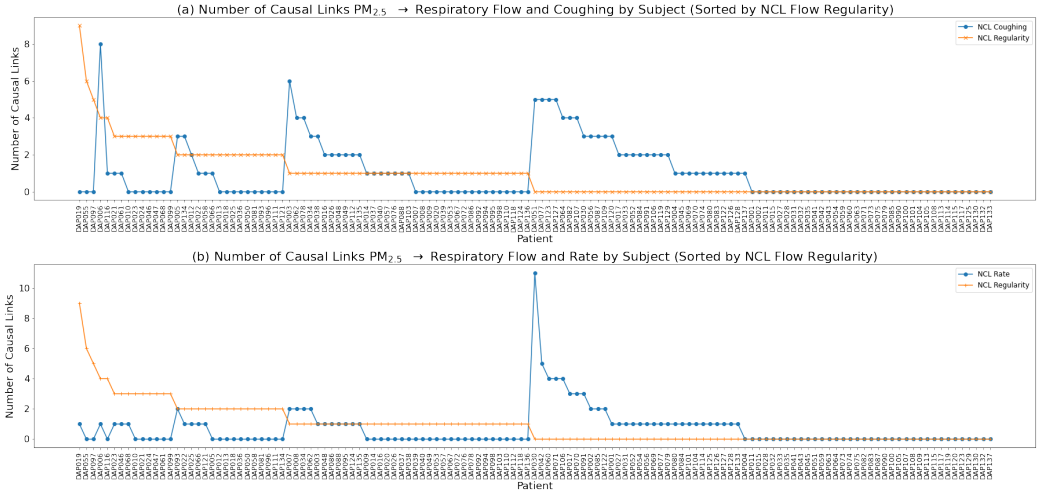


Figure 5.8: Comparison of Number of Causal Links (NCL) $PM_{2.5} \rightarrow$ Respiratory Effects. Coughing and Respiratory Rate Results Taken From [69]

Figure 5.8 compares the number of causal links, aggregated across both studies, from $PM_{2.5}$ to respiratory flow versus respiratory rate and coughing for each patient. An interesting observation can be made between the respiratory rate and respiratory flow results. All patients with a high number of causal links (three or more) to respiratory rate have zero causal links to respiratory flow. However, this is not a consistent pattern. Patients that have one causal link to respiratory flow also have similar variability in causal links to respiratory rate as patients with more than one causal link to respiratory flow. Therefore this negative correlation cannot be generalised across the dataset. Similarly, comparing the results of respiratory flow and coughing in Figure 5.8, it is even more evident that there is no consistent trend among the patients and the respiratory outcomes. This suggests that the respiratory outcomes of coughing or respiratory rate, and changes in the regularity of respiratory flow are sufficiently distinct to not be highly correlated. This also shows that the different respiratory outcomes are influenced by different physiological mechanisms. While rapid-shallow breathing is typically caused by stimulation of the Juxtacapillary-receptors in the aveolar walls [86], coughing and changes in respiratory rate are commonly induced by stimulation of the rapidly-adapting-receptors in the airway epithelium [10]. $PM_{2.5}$ therefore likely irritates a variety of receptors across patients in distinct ways, explaining this lack of a clear trend.

5.6 Sleep Covariate

Breathing patterns are more rapid and shallow for healthy patients during sleep stages than when awake [17]. Asthmatic patients also experience nocturnal asthmatic symptoms at different intensity and frequency than during the day due to normal physiologic change occurring at night [19]. Even hormonal changes at night may induce more severe asthmatic symptoms [27]. Therefore, to ensure the validity of this study and isolate the causal effect of $PM_{2.5}$ on the regularity of respiratory flow, sleep must be controlled for. Importantly, the activity level does not account for sleep episodes, and therefore a distinct variable had to be established for this.

However, this is conceptually a difficult task. Related research, including that on the impact of $PM_{2.5}$ on respiratory rate do not control for this variable [5]. Instead, the feature-set for sleep-wake classification developed in [78] was used. It consisted of generating the features in Table 5.1 from the raw RESpeck data, adding them to the fusion output in Table 3.2 and using them to predict the probability of sleeping.

The respiratory rate variability (RRV) is a frequency-domain feature that distinguishes linear, stationary and periodic breathing signals when asleep from those awake [78]. It is computed by taking the Fast Fourier Transform (FFT) of the breathing signal in expiration and applying Equation 5.1 from [36].

$$RRV = (1 - \frac{H1}{DC}) \times 100\% \tag{5.1}$$

where:

- $H1$ is the averaged amplitude of first harmonic peak over a 30-second window,
- DC is the averaged zero frequency amplitude over a 30-second window.

Field	Description
BR_md	Median breathing rate in 30 second window around breath
BR_mean	Mean breathing rate in 30 second window around breath
BR_std	Standard deviation of breathing rate in 30 second window
BR_coef	Coefficient of BR_std / BR_mean for 30 second window
AL_md	Median activity level in 30 second window around breath
AL_mean	Mean activity level in 30 second window around breath
AL_std	Standard deviation of activity level in 30 second window around breath
RRV	Respiratory rate variability
RRV3MA	Average of 3 RRV nearest neighbours

Table 5.1: Additional Sleeping Features Computed From Raw RESpeck Data

Having this feature set, transfer learning was attempted using the model in [78] trained on labeled data. Unsupervised learning with PCA and clustering was also tried. However, these methods proved ineffective. The dataset used in this investigation was vastly different from that in [78] and the model was unable to generalise. Meanwhile the unsupervised learning failed to converge to reasonable estimates of sleep probabilities. It predicted wakefulness throughout the night and sleep during the early hours of the day. To combat this, semi-supervised learning on a per-patient basis was done using the features in Table 5.1. The patients were assumed to be asleep between 2am and 6am and awake between 8am and 12pm. Noting that the patients were adolescents, these are reasonable assumptions given school hours. Subsequently, Scikit-learn's RandomForest [64] with 100 estimators was trained for each patient on those hours, and applied to the remaining hours' data. Its predicted probability of sleep was then combined into the PCMCI+ investigation setup, explained in Section 5.2, under the following assumptions:

- May have a contemporaneous and lagged link with activity level and flow regularity as the cause.
- Can be caused by a lag of itself, temperature, humidity, as evidenced in [41]

The PCMCI+ algorithm was then run on the expanded variable set with the results summarized in Figure 5.9 for the 1-hour study and Figure 5.10 for the 8-hour study. Both figures show the changes in causal links exhibited by patients when adding the sleeping parameter. The raw results illustrating the statistical significance of $PM_{2.5}$ to respiratory flow causal links across different time lags for each patient with the sleep parameter are depicted in Appendix C.3.1. The remaining PCMCI+ results are then outlined in Appendix C.3.

Figure 5.9 shows that despite adding the sleeping parameter to the 1-hour study, patients predominantly uphold their causal relationships between $PM_{2.5}$ and respiratory flow. 85.93% of patients with no causal links between $PM_{2.5}$ and respiratory flow before adding the sleep parameter, still had no causal links after its addition. Similarly, 74.67% of patients that had at least one causal link from $PM_{2.5}$ to respiratory flow before adding the sleep parameter, maintained at least one such link after its addition. These findings suggest that the influence of $PM_{2.5}$ on respiratory flow, where present, appears relatively

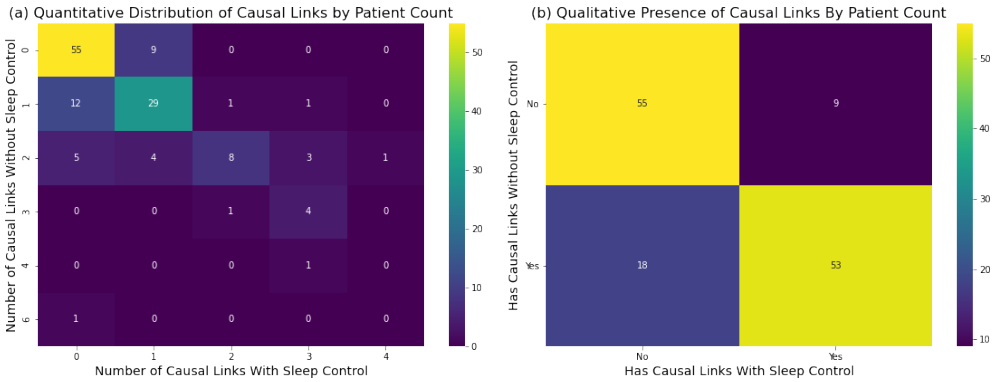


Figure 5.9: Distribution of $PM_{2.5} \rightarrow$ Respiratory Flow Causal Links With and Without Sleep Control For the 1H Study - Measured As The Number Of Patients

unaffected by sleep patterns for the majority of asthmatic patients. Consequently, for most patients demonstrating this causal relationship, the impact of $PM_{2.5}$ on respiratory flow persists across states of sleep and wakefulness. This matches expectations from [71], which associated $PM_{2.5}$ exposure with sleep-disordered breathing, showing its impact on respiratory functions even when asleep.

However, Figure 5.9 also shows that 18 patients who initially showed a causal link between $PM_{2.5}$ and respiratory flow no longer exhibited this relation after sleep was accounted for. Furthermore, adding the sleep parameter led to 9 new instances of a causal relationship being established. These findings demonstrate that while the impact of $PM_{2.5}$ on respiratory flow persists across states of sleep and wakefulness for the majority of patients, this is not applicable to all asthma patients. It thereby again highlights the heterogeneity in the patients' respiratory functions and their distinct respiratory responses to $PM_{2.5}$ exposure.

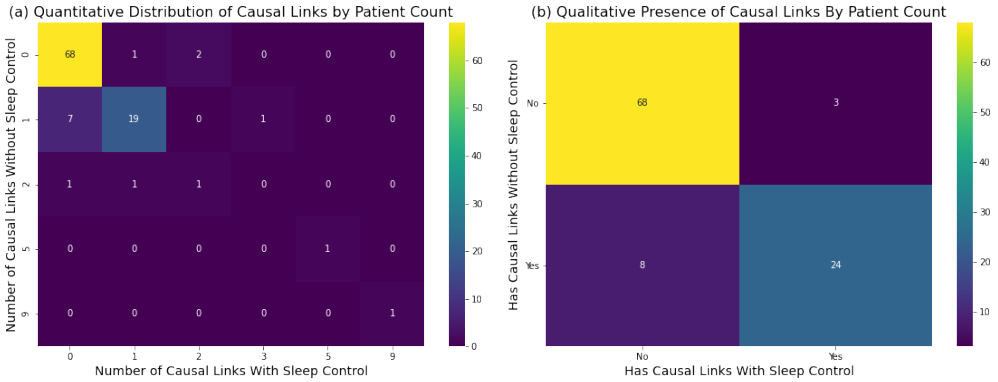


Figure 5.10: Distribution of $PM_{2.5} \rightarrow$ Respiratory Flow Causal Links With and Without Sleep Control For the 8H Study - Measured As The Number Of Patients

Figure 5.10 further supports the results and conclusions from the 1-hour study, in the 8-hour study. 95.77% of patients with no causal links between $PM_{2.5}$ and respiratory flow before adding the sleep parameter, still had no causal links after its addition. Similarly, 75.0% of patients that had at least one causal relation between $PM_{2.5}$ and

respiratory flow before adding the sleep parameter, maintained at least one such link after its addition. In fact, 85.43% of patients had the same number of causal links from $PM_{2.5}$ to respiratory flow after adding the sleep parameter as before. This further supports the conclusion that the causal relationship between $PM_{2.5}$ and respiratory flow is largely unaffected by sleep in most asthmatic patients.

While the sleep detection approach used in this section worked well, it is not optimal. Further work should be done to better classify episodes of sleep. Statistical models have shown to be effective in this regard [28], and more tuning can be done in this domain to optimize the results. Therefore, noting that the method used for sleep classification here cannot be guaranteed to be accurate because of the assumptions it makes to label training data, the remaining results in this report use the original PCMCI+ results without the sleep parameter. This ensures that the conclusions made are accurate.

5.7 Causal Results For Other Factors

Several factors beyond just $PM_{2.5}$ can impact the respiratory activities of asthmatic patients. Many of these are controlled for in the PCMCI+ algorithm in Section 5.2. To ensure a comprehensive analysis, the causal impact of other stimuli on respiratory flow is analysed in this section and contrasted to $PM_{2.5}$. The raw results showing the causal links from each cause to respiratory flow are presented in Appendix C.2, with selected insights presented in this section.

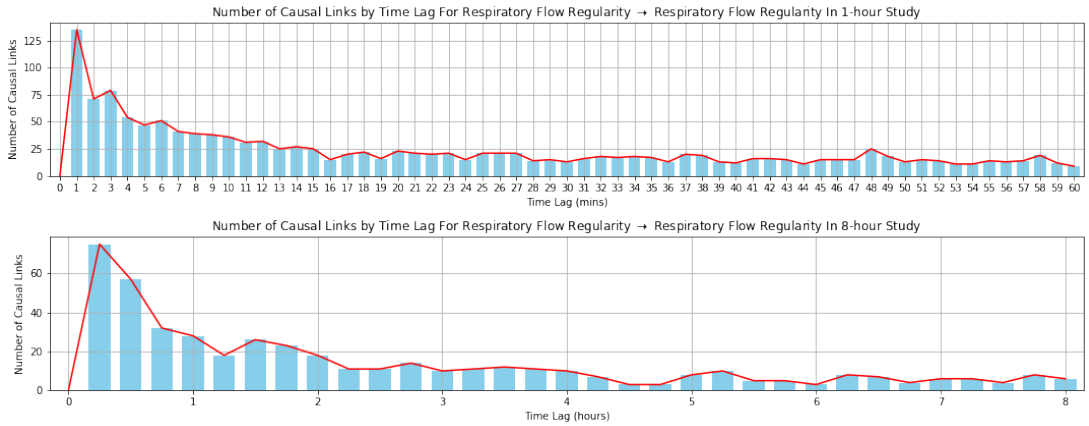


Figure 5.11: Number of Causal Links Respiratory Flow → Respiratory Flow By Time Lag In 1-hour (top) and 8-hour (bottom) Studies

Intuitively, it makes sense that our current respiration impacts the way we breathe in the nearest future. However, there is no clear understanding of how long current breathing will influence future respiration. Figure 5.11 shows the distribution of causal links from previous respiratory flow to current respiratory flow across time lags for both studies. In both studies, the number of causal links decays rather exponentially with time lag. This exponential decay of links was also found in the respiratory rate results from [57, 69]. This means that the frequency at which current breathing patterns impact future respiration diminishes exponentially over time. A similar analysis of temperature's, humidity's and activity level's causal links across time lags is in Appendix C.2.2.

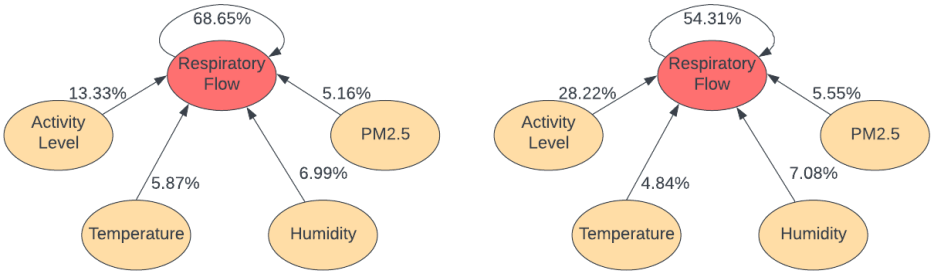


Figure 5.12: Proportion of Total Causal Links To Respiratory Flow By Cause In 1-hour Study (Left) and 8-hour Study (Right)

Meanwhile, Figure 5.12 shows the aggregated proportion of total causal links to respiratory flow by cause. As expected respiratory flow to respiratory flow has the most causal links. Meanwhile, activity level has the second most causal links, with the environmental stimuli having the least. In the 1-hour study, $PM_{2.5}$ has the least causal links, while in the 8-hour study it has marginally more than temperature only. Overall, temperature, humidity and $PM_{2.5}$ have comparable numbers of causal links. It is interesting that the environmental stimuli have substantially less causal links to respiratory flow than activity level and previous respiratory flow, in both studies. Appendix C.2.1 looks into this in more detail, and provides a comparison of the effects of environmental stimuli on respiratory flow.

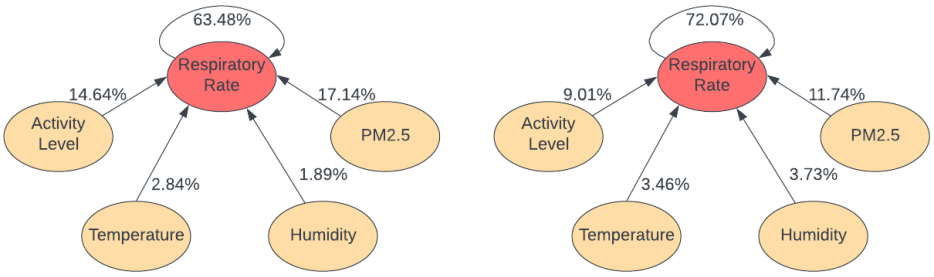


Figure 5.13: Proportion of Total Causal Links To Respiratory Rate By Cause In 1-hour Study (Left) and 8-hour Study (Right). Raw Results Taken From [57] Through [69]

Figure 5.13 uses results from [69, 57] to contrast the findings in Figure 5.12 against related work on the impact of $PM_{2.5}$ on respiratory rate. It shows the proportion of total causal links to respiratory rate by cause. In both the respiratory flow and respiratory rate projects, it is observed that previous respiratory activity has the most causal links to current respiratory activity. However, $PM_{2.5}$ exhibits a much greater proportion of links to respiratory rate compared to its proportion of links to respiratory flow (in both the 1-hour and 8-hour studies). Furthermore, unlike in the respiratory flow results, $PM_{2.5}$ has the second largest proportion of links to respiratory rate, greater than activity level even. There are also generally more links and patients with links from $PM_{2.5}$ to respiratory rate than from $PM_{2.5}$ to respiratory flow, in both studies. This suggests that respiratory rate is more sensitive than respiratory flow to $PM_{2.5}$.

Chapter 6

Causal Effects

6.1 Causal Effect Estimation

In Chapter 5 we have identified causal links between factors like $PM_{2.5}$ and changes in the regularity of respiratory flow. To better understand these causal relations, the causal effect estimation algorithm from Runge et al. [68] is used to compute the strengths of the causal links. Doing this is crucial to understand how $PM_{2.5}$ impacts respiratory flow.

The effect estimation algorithm uses the PCMCI+ results to construct a causal network graph. This graph has nodes for each variable and time lag. These nodes are connected by edges representing the established causal links from PCMCI+. It then applies linear regression on the corresponding time series data to compute the coefficients of the causal links (edges). This is known as the direct causal effect. It then back propagates across the causal graph to estimate the indirect causal effects. This is explained in more detail in Algorithm 5. The implemented algorithm itself is taken and adapted from [57]. In rare cases, the instability of linear regression [22] in Algorithm 5 resulted in unreasonably large causal effect strengths. These were controlled for by setting the maximum absolute effect strength to 1 and capping the rare larger values. Importantly, this infrequent adjustment ensured consistency in the analysis and facilitated the interpretation of the results without altering the overall trends.

Although alternative methodologies to causal effect estimation exist [4] like the Auto-G-Computation algorithm [77], they are less compatible with this task. They often have parametric assumptions that require homoscedasticity in the data. These assumptions cannot be made in this task, where sensor data is used that varies greatly across patients, particularly in respiratory observations. Therefore, Algorithm 5 from [68] is used, which is tailored to observational time-series data and does not make those assumptions.

6.2 Causal Effect Comparison

The causal effect estimation from algorithm 5 returns two types of effects: direct and indirect effects. The direct effect of one variable on another is the influence it exerts on the other variable without mediation by other variables. Meanwhile an indirect effect is

Algorithm 5 Causal Effect Estimation with Linear Regression Weights

Input: Time series data $\mathbf{X}$, Maximum lag $\tau_{\max}$, Significance threshold α_{PC}
Output: Causal Effects **CE**, Directed Causal Effects **DCE**

function CALCULATECAUSALEFFECTS($\mathbf{X}, \tau_{\max}, \alpha_{\text{PC}}$)
  $\mathcal{G} \leftarrow$ PCMCI+($\mathbf{X}$, $\tau_{\max}$, α_{PC})
  Initialize $\mathbf{W}$ matrix for edge weights
  for each variable X^j in $\mathbf{X}$ **do**
    Let $\mathcal{M}$ be a Linear Regression Model
    Let $\mathcal{D} \leftarrow \{X^i_{t-\tau}$ for all $X^i \in \mathcal{G}$ for all $0 \leq \tau \leq \tau_{\max} : (X^i_{t-\tau}, X^j_t) \in \mathcal{G}\}$
    Fit $\mathcal{M}$ to predict X^j_t from $\mathcal{D}$
    for each edge $(X^i_{t-\tau}, X^j_t)$ in $\mathcal{G}$ **do**
      $w \leftarrow$ standardized regression coefficient of $X^i_{t-\tau}$ in $\mathcal{M}$
      $\mathbf{W}(X^i_{t-\tau}, X^j_t) \leftarrow w$
    end for
  end for
  Modify edges in $\mathcal{G}$ to have corresponding value in $\mathbf{W}$
  Initialize **CE** and **DCE** matrices
  for each variable X^j in $\mathbf{X}$ **do**
    Define target as X^j_0
    Compute causal effects for X^j_0 using backpropagation in $\mathcal{G}$
    Compute direct causal effects for X^j_0 from direct edges to X^j_0 in $\mathcal{G}$
    Store results in **CE** and **DCE**
  end for
  return CE, **DCE**
end function

mediated through one or more additional variables. This accounts for situations where, for example, $PM_{2.5}$ effects temperature which then effects respiratory flow. Typically, there are more indirect than direct effects. This is because they accumulate through chained links across multiple time lags. This indeed is reflected in the results in Figure 6.1 and Figure 6.2.

Figures 6.1 and 6.2 show the number of direct and indirect causal effects to respiratory flow by cause for each study. The causal effects are split into positive and negative effects. A positive effect means that an increase in the cause variable induces a more regular respiratory flow pattern in the patient. Otherwise, a negative effect means that an increase in the cause variable prompts a more irregular respiratory flow in the patient. Overall, the trends in number of causal effects are consistent across both studies. This consistency is promising and expected, as it suggests a reliable pattern in how different variables influence respiratory flow across different time-frames.

A general observation is that, for all causes, the number of causal effects in Figures 6.1 and 6.2 is less than the number of causal links established in Chapter 5. This is because certain statistically significant links in the causal discovery model may translate into statistically insignificant causal effects when considering their magnitude and direction

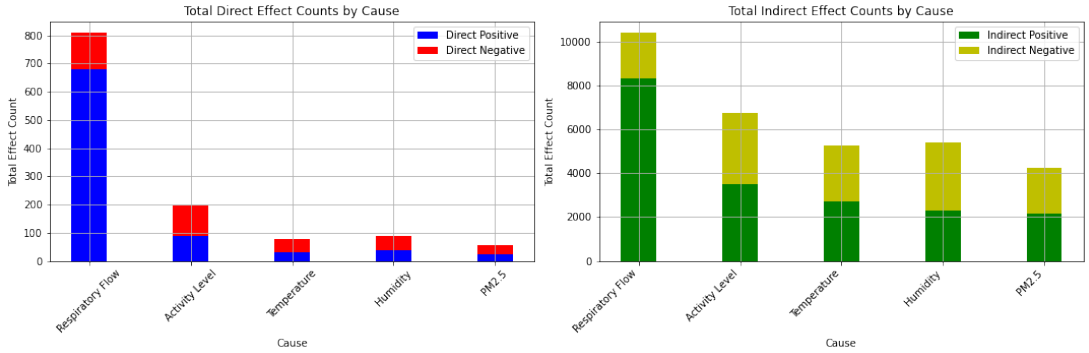


Figure 6.1: Number of Causal Effects To Respiratory Flow, Across All Patients, By Cause In 1-hour Study

of influence. There are also more effects for each cause in the 1-hour study than the 8-hour study. Initially, this might suggest that the selected variables impact respiratory flow more frequently within the first hour time-frame. Yet this is not necessarily the case. For each patient, each variable can have up to 60 effects in the 1-hour study, but only up to 32 effects in the 8-hour study. The 8-hour study also ran on less patients than the 1-hour study. These differences in the upper bound of causal effects may therefore explain this observation.

In both studies, preceding regularity of respiratory flow has the most direct and indirect causal effects on the current regularity of respiratory flow. This is especially visible in the 1-hour study where preceding respiratory flow influences the current respiratory flow with 4.13 times the number of direct effects compared to the cause with the second highest amount of effects. These results makes sense given breathing is a continuous process where the way a person currently breathes impacts their subsequent breaths. In both studies, activity level has the second most direct and indirect causal effects to current respiratory flow. This too matches expectations given the almost immediate respiratory responses humans have to changes in physical activity intensity [45].

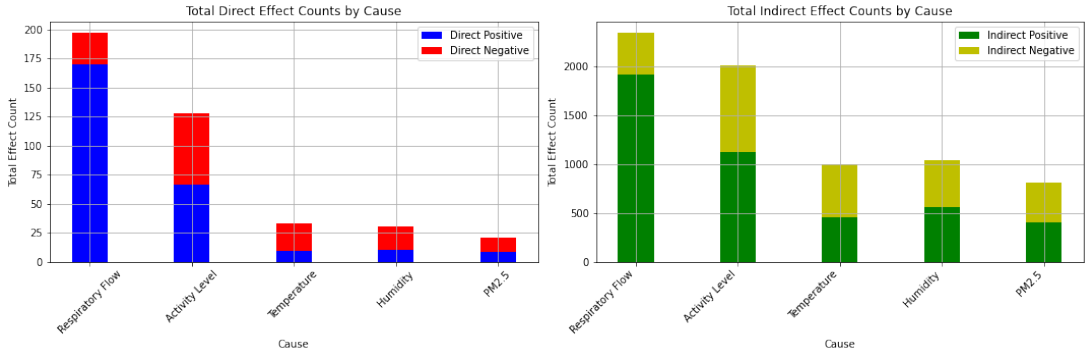


Figure 6.2: Number of Causal Effects To Respiratory Flow, Across All Patients, By Cause In 8-hour Study

Further analysing the results, $PM_{2.5}$ has the least direct and indirect causal effects to respiratory flow in both the 1-hour and 8-hour studies. This means that $PM_{2.5}$ impacts

respiratory flow less frequently than the other factors of temperature, humidity, previous respiratory flow and activity level. $PM_{2.5}$ also has a comparable split between positive and negative effects on respiratory flow in both the studies. This can be seen in both the direct and indirect effects in Figures 6.1 and 6.2. This result once again underscores the variability in every patient's respiratory response to environmental factors.

Although a positive effect may seem counter-intuitive, as it implies exposure to increased $PM_{2.5}$ elicits a more regular respiratory flow: this result is not unreasonable. While studies have shown warmer ambient temperature to decrease lung function in the general population [15], higher temperatures were also found to decrease asthma severity and symptoms in certain patients [16]. The field of asthmatic responses to environmental stimuli is greatly different to that of healthy patients and remains largely unexplored. Therefore, with this in mind and in consultation with Dr Gordon Drummond on these findings, such results are plausible.

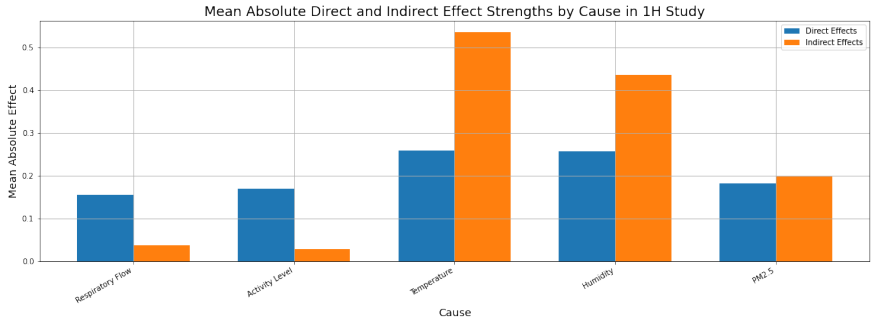


Figure 6.3: Mean Absolute Causal Effect Strength To Respiratory Flow In 1-hour Study

Figures 6.3 and 6.4 show the mean absolute effect strength from each cause to respiratory flow in the 1-hour and 8-hour studies respectively. The trends observed in these figures are again consistent across both studies. Temperature and humidity have the largest direct and indirect effect strengths. They are followed by $PM_{2.5}$, which has a mean absolute effect strength greater than previous respiratory flow and activity level. These results show that although environmental factors ($PM_{2.5}$, temperature, humidity) have significantly less causal effects to respiratory flow than activity level and previous respiratory flow (Figures 6.1 and 6.2), on average their causal effects are much stronger. Therefore, when patients' respiratory flow is influenced by $PM_{2.5}$, they experience more pronounced changes in respiration compared to impacts stemming from non-environmental factors. Initially, it may seem surprising that activity level has a relatively small impact on respiratory flow. However, the data for this investigation is only collected during stationary activities. These activities are typically of low intensity and therefore it makes sense that they induce a relatively small causal effect on respiratory flow.

The Figures also show that the mean direct effects on respiratory flow are significantly larger in the 8-hour study compared to the 1-hour study for all causes. Therefore, although fewer patients exhibit causal links from $PM_{2.5}$ to respiratory flow at 15-minute resolution over 8 hours, the induced effects are notably stronger than those affecting 1-minute respiratory flow periods within the first hour after exposure. This highlights the delayed, yet intensified, impact that $PM_{2.5}$ can have on respiration.

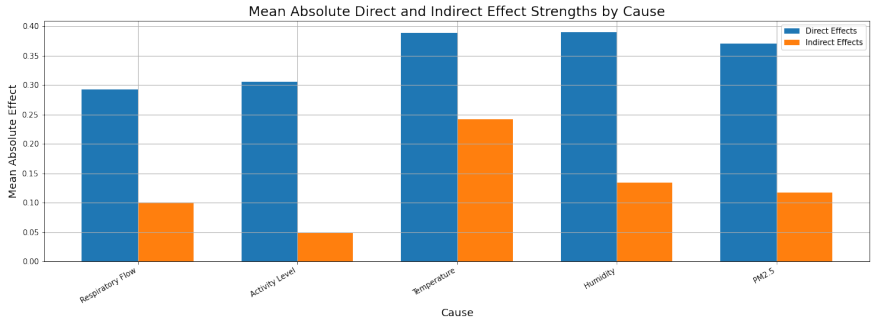


Figure 6.4: Mean Absolute Causal Effect Strength To Respiratory Flow In 8-hour Study

6.3 Temporal Causal Effect Analysis

The results in Section 6.2 offer valuable insights into the varying causal effects produced by different variables. Those insights, however, are aggregated across time, overlooking potential variations in causal effects over different time lags or months. This section therefore focuses on investigating the patterns in causal effects across time.

Figure 6.5 shows the mean and range of the strength of $PM_{2.5}$ effects on respiratory flow across time lags in the 1-hour study. The direct effect strengths are observed to be highly variable across the sixty minutes with sporadic peaks and no consistent pattern. The strength directions change from positive to negative rapidly and have no clear pattern in magnitude. This is likely a side-effect of aggregating the distinct causal responses of each patient together. Each patient has responses of different magnitude and direction that are spread across the time lags. This likely explains the high variability in strengths across lags. A similar trend can also be seen in the distribution of direct effect strengths across the 8-hour lags in Figure 6.6. These strengths are observed to be even more volatile than those in the 1-hour study.

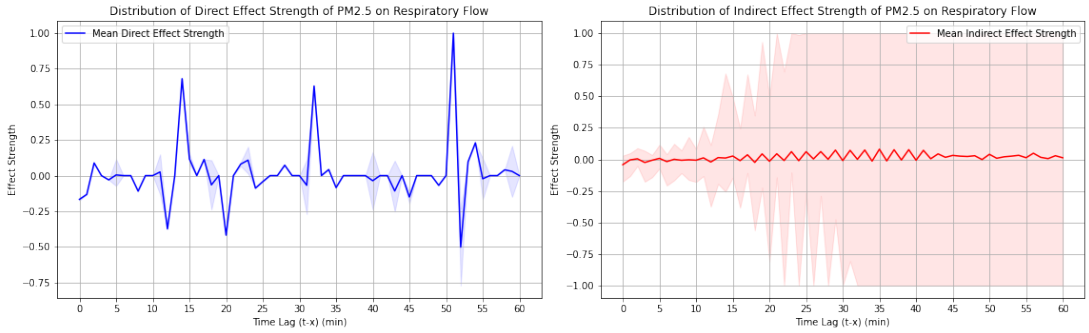


Figure 6.5: Distribution of $PM_{2.5} \rightarrow$ Respiratory Flow Causal Effect Strength Across Time Lags in 1-hour Study

In turn, the indirect effect strengths consistently have zero-like means in both studies for all time lags, with occasional deviations. It is also observed in both Figures 6.5 and 6.6, that the variance in the distribution of indirect effect strengths increases with the time lags. This is explained by the cumulative nature of indirect effects in algorithm 5. For larger time lags, the algorithm must back-propagate through growing chains of

direct causal effects. This therefore results in the observed divergent nature of indirect effects in both figures and is depicted by the quickly intensifying positive and negative indirect effect strengths.

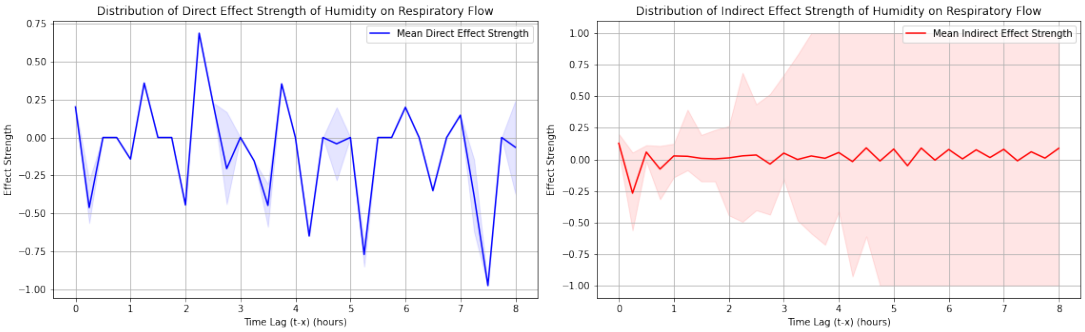


Figure 6.6: Distribution of $PM_{2.5} \rightarrow$ Respiratory Flow Causal Effect Strength Across Time Lags in 8-hour Study

The next step of this analysis is to look at how effect strength varies across explicit time periods throughout the year. Recent studies demonstrated that $PM_{2.5}$ concentrations exhibit monthly and seasonal patterns [35, 43]. A hypothesis therefore exists that the causal effect of $PM_{2.5}$ on respiratory flow changes across seasons or months. To investigate this we look at the distributions of direct effects across months, noting that the indirect effects are too unstable to analyse. Figure 6.7 depicts the mean and the range of the direct effect strengths from $PM_{2.5}$ to respiratory flow in each month of the year in the 1-hour study. It is observed that the direct effects typically vary greatly in direction and magnitude within the same month and that there are no significant outlier months. The months of March, April and July have causal effects of predominantly negative direction. The remaining months have a larger range of causal effect strengths, spanning both positive and negative values. Meanwhile, the months of March, June and July have the narrowest range in effect strength magnitude, having substantially smaller causal effect strength magnitudes than the effects in other months. September and April generally have the strongest effect strength magnitudes. Interestingly, October has no direct causal effects in the 1-hour study despite having causal links in Figure 5.7. This suggests that the short-term causal relations in October are not strong enough to have a significant impact on respiratory flow. Meanwhile, May has no causal effects because no data was collected in that month.

Nonetheless, an interesting trend is that the months near summer, particularly June and July but also March and August, typically have the least variable effect strengths and have predominantly smaller magnitudes compared to the other months. A study by Singh et al. showed that the $PM_{2.5}$ concentrations in Delhi, India are typically substantially smaller in those months than in other months like those of winter [72]. This could suggest a correlation between higher $PM_{2.5}$ concentrations and more variable, often stronger, short-term causal effects from $PM_{2.5}$ to respiratory flow. Although April is an outlier to this trend having both a strong mean effect and a large effect range, it is skewed by a single patient having one strong effect, while all other patients in April conform to the trend. This patient could very likely just be an outlier to this observed

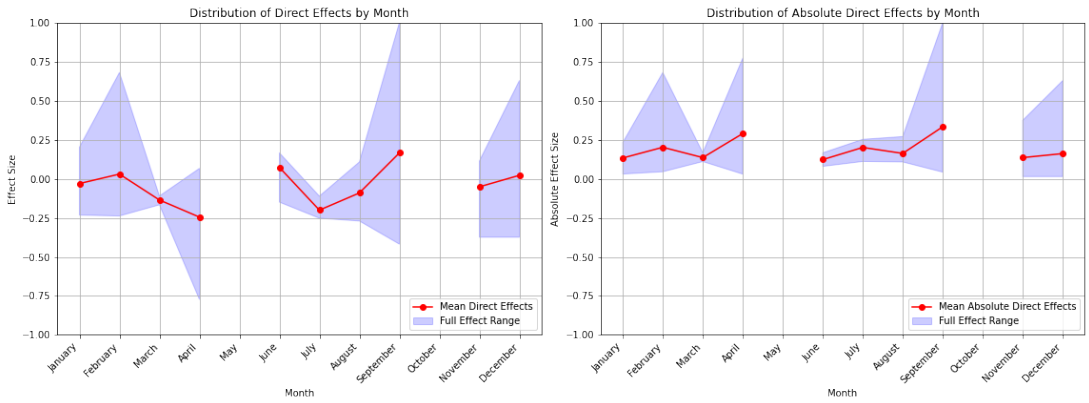


Figure 6.7: Range and Mean of Direct $PM_{2.5} \rightarrow$ Respiratory Flow Causal Effect Strength Across Months in 1-hour Study

pattern. Nonetheless, further research, with more data distributed equally across the months of the year would be needed to decisively conclude this association.

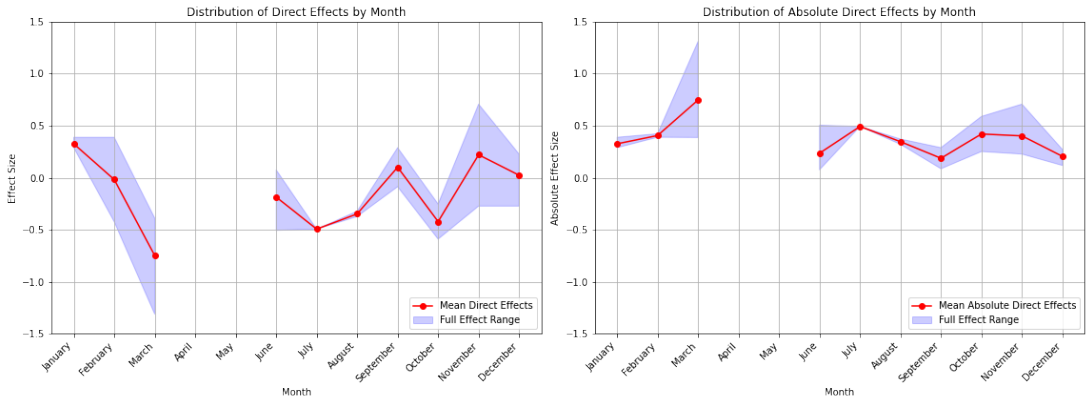


Figure 6.8: Range and Mean of Direct $PM_{2.5} \rightarrow$ Respiratory Flow Causal Effect Strength Across Months in 8-hour Study

Figure 6.8 shows the mean and the range of the direct effect strengths from $PM_{2.5}$ to respiratory flow in each month of the year for the 8-hour study. In this study, the effect strengths vary significantly less intra-month than in the 1-hour study. However, this could just be a reflection of the much smaller number of causal effects in the 8-hour study. Nonetheless, March and July again, but also August and October, have predominantly negative, but also some of the strongest, causal effects. This demonstrates that increased $PM_{2.5}$ typically makes respiratory flow significantly more irregular in those months. Finally, Appendix D.1 shows the mean and the range of the direct effect strengths from $PM_{2.5}$ to respiratory rate across months using results from [69]. Contrasting these results against respiratory flow, $PM_{2.5}$ typically induces a more variable in magnitude and direction effect on respiratory rate than respiratory flow. However, this could be because there are generally more causal effects from $PM_{2.5}$ to respiratory rate than respiratory flow. An interesting observation though is that in the 8-hour studies, in October, $PM_{2.5}$ induces strong negative-only causal effects on both respiratory rate and flow.

Chapter 7

Conclusions

7.1 Discussion

In this investigation, a comprehensive analysis of the impact of $PM_{2.5}$ on respiratory flow in asthmatic adolescents was performed. For this, RESpeck respiratory and physical activity data and AIRSpeck environmental data collected on 137 patients in the DAPHNE study was used. The raw sensor data was firstly processed to accurately capture breaths from the breathing signal. This was done using the ”gold standard” method that employs sliding windows of signal thresholds to accurately handle sensor noise. Next, the localized mean respiratory rate and tidal volume approximations were extracted as features for each breath. These tidal volume approximations were derived by calculating both the peak respiratory flow and the area under the breathing signal curve. Then, the computed breath data and AIRSpeck data were fused together, with outliers carefully identified and removed. To overcome the assumptions made by linear interpolation in related research, multivariate imputation by chained equations was applied and tested on missing environmental data to enhance the dataset's completeness.

Using the features computed for each breath, a medically-informed metric was then developed in consultation with a medical professional. It measured the regularity of respiratory flow of each breath. This metric employed principal component analysis to evaluate the deviation of each breath's features from the principal components. It thereby successfully identified breaths that are more rapid-shallow and have higher frequency as more irregular. It is non-invasive in nature and is the first metric ever designed to measure the quality of respiratory flow.

Next, a causal analysis between $PM_{2.5}$ and respiratory flow was performed. To motivate the analysis, $PM_{2.5}$ variables were first shown to improve the mean adjusted R^2 of regression models predicting the regularity of respiratory flow. Then, PCMCI+ was used to examine the causal relations between $PM_{2.5}$ and respiratory flow, while accounting for temperature, humidity and activity level. PCMCI+ is a method for discovering lagged and contemporaneous causal links in time series data, making it suitable for this task. To analyse both short-term and longer-term causal relations, the method was applied with max lag of 1-hour on 1-minute resolution data and also with max lag of 8-hours on 15-minute resolution data.

The results showed that at a p-value threshold of 0.02, 52.29% of patients exhibit causal links from $PM_{2.5}$ to respiratory flow in the 1-hour study and 31.07% of patients in the 8-hour study. In the 1-hour study, patients that exhibit causal links are on average causally impacted at 1.549 time lags while in the 8-hour study this value decreases to 1.469. These findings indicate that $PM_{2.5}$ has a significant causal relation with respiratory flow that is prevalent in a majority of asthmatic patients. Next, a survival analysis showed that although the causal links in both studies are distributed rather uniformly across the time lags, the incidence rate of a patient having their first causal link diminishes with time. In the 1-hour study patients are substantially more likely to have a causal link from $PM_{2.5}$ to respiratory flow in the first 35-minutes than in minutes 35-60. The same is observed in the 8-hour study where the incidence rate is significantly higher in the time lags of the first 2 hours than hours 6 to 8. This lets us conclude that patients are more likely to have their first respiratory flow response to $PM_{2.5}$ in the earlier phases of exposure. Finally, the incidence of causal links in each patient is compared between respiratory flow and other respiratory conditions. The results showed that there are no consistent trends in number of causal links. This indicates that the $PM_{2.5}$ induced changes in coughing or respiratory rate are distinct and influenced by different physiological mechanisms than those in respiratory flow.

Next, to truly isolate the causal links between $PM_{2.5}$ and respiratory flow, the causal analysis was repeated with an additional variable of sleep probability. This was motivated by the fact that breathing patterns and asthma symptoms vary between sleep and wakefulness states. The sleeping probability was predicted using semi-supervised learning on a set of features derived from the breathing rates and activity levels. In the 1-hour study, 85.93% of patients without causal links and 74.67% with links before adding sleep probability maintained their link status afterward. In the 8-hour study, these values were 95.77% and 75.0%, respectively, with 85.43% of patients retaining the exact same number of links. These results show that sleep has a minimal impact on the relationship between $PM_{2.5}$ and respiratory flow, with the majority of asthmatic patients' responses to $PM_{2.5}$ being consistent across states of sleep and wakefulness.

Finally, the established causal relations between $PM_{2.5}$ and respiratory flow were quantified. This was achieved using a causal effect estimation algorithm, which constructed a causal network using the PCMCI+ results. It then employed linear regression on time series data to compute the causal effects. The results showed that although $PM_{2.5}$ had the least causal effects to respiratory flow among the factors, they were on average substantially stronger than the effects of activity level and previous respiratory flow but weaker than the effects of temperature and humidity. This was consistent for both direct and indirect effects in both the 1-hour and 8-hour studies. Therefore, it is concluded that $PM_{2.5}$ has a strong causal effect on the regularity of respiratory flow and induces substantial changes in the way asthmatics breath. Overall, the less frequent, longer-term effects in the 8-hour study were stronger than those in the 1-hour study. The results also showed that the causal effects of $PM_{2.5}$ on respiratory flow fluctuated in both magnitude and sign across different time lags. This demonstrated that there is no uniform respiratory flow response to $PM_{2.5}$ exposure across patients. The results also hinted a monthly association between higher $PM_{2.5}$ concentrations and more variable, stronger, effect strengths. However, further research is needed to conclude this decisively.

7.2 Limitations and Future Work

One limitation of the developed metric to quantify the regularity of respiratory flow is that it is not very generalisable. It makes use of the distributions of breath features to identify rapid-shallow and high frequency breaths. However, different methods of capturing breaths or computing the breathing signal will likely have different distributions of breath features for which the metric might not work as well. Future work can therefore build on this metric and utilise alternative unsupervised learning techniques to make it usable on any respiratory dataset and method of capturing breaths.

Another limitation of the existing approach is that the metric currently quantifies the regularity of respiratory flow for each breath. These are then averaged to obtain a score for the regularity of respiratory flow for 1-minute and 15-minute respiratory periods. This averaging process ensures that outlier breaths do not skew the regularity of respiratory flow assigned to these time periods. However, perhaps a more accurate approach would be to directly measure the regularity of respiratory flow for the entire interval rather than for each breath. This would allow the metric to consider how respiratory flow changes across successive breaths when determining the regularity score. Another way future work can do this, is by defining a custom scoring function that considers successive breaths and the rate of respiratory flow change. This function could then replace the mean function when aggregating regularity scores from individual breaths into these intervals.

Future work can also focus on improving the sleep episode classification. The approach used in this study is limited because it assumes that patients are awake at certain times and asleep in others to generate a training set for semi-supervised learning. Although these assumptions were kept at a minimum to ensure the results were accurate, further work can be done to improve this. Unsupervised learning techniques can be further experimented with to obtain better estimates of wakefulness and sleep episodes. In fact, instead of predicting the probability of sleep, future work could look into additionally identifying the stage of sleep a patient is in. Different sleep stages have been shown to affect respiration differently [88]. Doing this would therefore help better understand the impact of sleeping on respiratory flow and more accurately isolate the causal effects of $PM_{2.5}$ on respiratory flow from other confounding factors.

Moreover, future work can look into accounting for more potential confounding factors that may impact respiratory flow. These were outlined in Chapter 2 and while accounting for temperature, humidity and activity level is a good start, further research could expand these to better isolate the causal effects of $PM_{2.5}$ on respiratory flow from other confounding factors. Further work can also be done to understand the causal relation between other air pollutants and respiratory flow. These may include particular matter of broader or narrower diameter dimensions and different air pollutants like PM_{10}, PM_1 and nitrogen dioxide. All of these have been found to be correlated with exasperated respiratory conditions [74, 57, 31] and would therefore be worthwhile to investigate in the context of respiratory flow responses. Finally, more data can be collected and analysed to investigate possible seasonal associations between $PM_{2.5}$ concentration distributions and causal effect strength of $PM_{2.5}$ on respiratory flow.

Bibliography

[1] Bernard Hannhart Abraham B. Bohadana and Dan B. Teculescu. Nocturnal worsening of asthma and sleep-disordered breathing. *Journal of Asthma*, 39(2):85–100, 2002. PMID: 11990234.

[2] Kalolella Admirabilis, Beno. Sleeping position and reported night-time asthma symptoms and medication. *Pan African Medical Journal*, 24(59), 05 2016.

[3] Trompenaars MP Amanda, Van Roest PJ Aalt, and Vaessen-Verberne APH Anj. Dysfunctional breathing in children. *Journal of Pulmonology and Respiratory Research*, 4(1):001–005, May 2020.

[4] Weihua An, Roberson Beauvile, and Benjamin Rosche. Causal network analysis. *Annual Review of Sociology*, 48(1):23–41, 2022.

[5] Pablo Sedeno Andreu. A data-driven analysis of the impact of short-term exposure to air pollution on the respiratory rate of asthmatic adolescents. Master's thesis, School of Informatics, University of Edinburgh, 2020.

[6] D. K. Arvind, Janek Mann, Andrew Bates, and Konstantin Kotsev. The airspeck family of static and mobile wireless air quality monitors. In *2016 Euromicro Conference on Digital System Design (DSD)*, pages 207–214, 2016.

[7] Asthmaandlung. Peak flow, Nov 2022.

[8] Melissa J. Azur, Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1):40–49, Feb 2011.

[9] Andrew Bates, Martin J. Ling, Janek Mann, and D.K. Arvind. Respiratory rate and flow waveform estimation from tri-axial accelerometer data. In *2010 International Conference on Body Sensor Networks*, pages 144–150, 2010.

[10] Joshua E. Brinkman, Fadi Toro, and Sandeep Sharma. Physiology, respiratory drive, Jun 2023.

[11] J.R. Catterall, P.M.A. Calverley, V. Brezinova, N.J. Douglas, H.M. Brash, C.M. Shapiro, and D.C. Flenley. Irregular breathing and hypoxaemia during sleep in chronic stable asthma. *The Lancet*, 319(8267):301–304, Feb 1982.

[12] Linlin Chen, Huidan Deng, Hengmin Cui, Jing Fang, Zhicai Zuo, Junliang Deng,

Yinglun Li, Xun Wang, and Ling Zhao. Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget*, 9(6):7204–7218, Dec 2017.

[13] Yuxiong Chen, Dehui Kong, Jia Fu, Yongqiao Zhang, Yakun Zhao, Yanbo Liu, Zhen'ge Chang, Yijie Liu, Xiaole Liu, Kaifeng Xu, Chengyu Jiang, and Zhongjie Fan. Associations between ambient temperature and adult asthma hospitalizations in beijing, china: a time-stratified case-crossover study. *Respiratory Research*, 23, 02 2022.

[14] Chetna, Surendra K. Dhaka, Gagandeep Longiany, Vivek Panwar, Vinay Kumar, Shristy Malik, A.S. Rao, Narendra Singh, A.P. Dimri, Yutaka Matsumi, and et al. Trends and variability of pm2.5 at different time scales over delhi: Long-term analysis 2007–2021. *Aerosol and Air Quality Research*, 23(5):220191, 2023.

[15] Joseph M. Collaco, Lawrence J. Appel, John McGready, and Garry R. Cutting. The relationship of lung function with ambient temperature. *PLOS ONE*, 13(1):1–11, 01 2018.

[16] R J Delfino, R S Zeiger, J M Seltzer, D H Street, R M Matteucci, P R Anderson, and P Koutrakis. The effect of outdoor fungal spore concentrations on daily asthma severity. *Environmental Health Perspectives*, 105(6):622–635, Jun 1997.

[17] N J Douglas, D P White, C K Pickett, J V Weil, and C W Zwillich. Respiration during sleep in normal man. *Thorax*, 37(11):840–844, Nov 1982.

[18] Gordon B. Drummond, Darius Fischer, Margaret Lees, Andrew Bates, Janek Mann, and D.K. Arvind. Classifying signals from a wearable accelerometer device to measure respiratory rate. *ERJ Open Research*, 7(2), 2021.

[19] Carolyn M. D'Ambrosio and Vahid Mohsenin. Sleep in asthma. *Clinics in Chest Medicine*, 19(1):127–137, Mar 1998.

[20] Stefan Edginton, Dylan E. O'Sullivan, Will D. King, and M. Diane Lougheed. The effect of acute outdoor air pollution on peak expiratory flow in individuals with asthma: A systematic review and meta-analysis. *Environmental Research*, 192:110296, 2021.

[21] N. Eid, Ben Yandell, L. Howell, Martha E. Eddy, and S. Sheikh. Can peak expiratory flow predict airflow obstruction in children with asthma? *Pediatrics*, 105 2:354–8, 2000.

[22] Steven P. Ellis. Instability of least squares, least absolute deviation and least median of squares linear regression, with a comment by stephen portnoy and ivan mizera and a rejoinder by the author. *Statistical Science*, 13(4), Nov 1998.

[23] Elizabeth Emrath. The basics of ventilator waveforms. *Current Pediatrics Reports*, 9(1):11–19, Jan 2021.

[24] EPA. Particulate matter (pm) basics, Jul 2023.

[25] Centre for Speckled Computing. Pollution sensors help asthmatics' health, Apr 2023.

[26] J. Mann G. B. Drummond, A. Bates and D. K. Arvind. Validation of a new non-invasive automatic monitor of respiratory rate for postoperative subjects. In *British Journal of Anaesthesia*, pages 462–469, 2011.

[27] GAAPP. Asthma at night: Causes, symptoms, treatment and more information, Apr 2023.

[28] Maksym Gaiduk, Thomas Penzel, Juan Antonio Ortega, and Ralf Seepold. Automatic sleep stages classification using respiratory, heart rate and movement signals. *Physiological Measurement*, 39(12):124008, Dec 2018.

[29] Sermin Genc, Zeynep Zadeoglulari, Stefan H. Fuss, and Kursad Genc. The Adverse Effects of Air Pollution on the Nervous System. *Journal of Toxicology*, 2012:782462, February 2012. Publisher: Hindawi Publishing Corporation.

[30] Adel Ghorani-Azam, Bamdad Riahi-Zanjani, and Mahdi Balali-Mood. Effects of air pollution on human health and practical measures for prevention in Iran. *Journal of Research in Medical Sciences*, 21(1), 2016.

[31] J. Gillespie-Bennett, N. Pierse, K. Wickens, J. Crane, P. Howden-Chapman, , the Housing Heating, and Health Study Research Team. The respiratory health effects of nitrogen dioxide in children with asthma. *European Respiratory Journal*, 38(2):303–309, 2011.

[32] Sean A Gilman and Robert B Banzett. Physiologic changes and clinical correlates of advanced dyspnea. *Current Opinion in Supportive; Palliative Care*, 3(2):93–97, Jun 2010.

[33] Alexander F. Glick, Suzy Tomopoulos, Arthur H. Fierman, Anne Elixhauser, and Leonardo Trasande. Association between outdoor air pollution levels and inpatient outcomes in pediatric pneumonia hospitalizations, 2007 to 2008. *Academic Pediatrics*, 19(4):414–420, 2019.

[34] Paul Grossman, Monika Spoerle, and Frank Wilhelm. Reliability of respiratory tidal volume estimation by means of ambulatory inductive plethysmography. *Biomedical sciences instrumentation*, 42:193–8, 02 2006.

[35] Air Quality Expert Group. *Fine Particulate Matter (PM2.5) in the United Kingdom*. Defra - Department for Environment Food and Rural Affairs, 2012.

[36] Guillermo Gutierrez, Jeffrey Williams, Ghadah A. Alrehaili, Anna McLean, Ramin Pirouz, Richard Amdur, Vivek Jain, Jalil Ahari, Amandeep Bawa, and Shawn Kimbro. Respiratory rate variability in sleeping adults without obstructive sleep apnea. *Physiological Reports*, 4(17), Sep 2016.

[37] Wiebke Günther, Urmi Ninad, jonas Wahl, and Jakob Runge. Conditional independence testing with heteroskedastic data and applications to causal discovery, 2023.

[38] Sasha Hallett. Physiology, tidal volume. *National Library of Medicine*, May 2023.

[39] Don Hayes, Paul B. Collins, Mehdi Khosravi, Ruei-Lung Lin, and Lu-Yuan Lee. Bronchoconstriction triggered by breathing hot humid air in patients with asthma.

American Journal of Respiratory and Critical Care Medicine, 185(11):1190–1196, 2012. PMID: 22505744.

[40] Chest Heart and Stroke Scotland. Normal breathing. *RESPe*, 2023.

[41] Eric Horvath. Study links air pollution, heat, noise and more to reduced sleep. *Penn Medicine*, Apr 2023.

[42] Daisy J.A. Janssen, Emiel F.M. Wouters, and Martijn A. Spruit. Psychosocial consequences of living with breathlessness due to advanced disease. *Current Opinion in Supportive and Palliative Care*, 9(3):232–237, Sep 2015.

[43] Wei Ji, Yong Wang, and Dafang Zhuang. Spatial distribution differences in pm2.5 concentration between heating and non-heating seasons in beijing, china. *Environmental Pollution*, 248:574–583, 2019.

[44] Xu-Qin Jiang, Xiao-Dong Mei, and Di Feng. Air pollution and chronic airway diseases: what should people know and do? *Journal of Thoracic Disease*, 8(1), 2015.

[45] European Respiratory Journal. Your lungs and exercise. *Breathe*, 12(1):97–100, 2016.

[46] Ryo Kamoi and Kei Kobayashi. Why is the mahalanobis distance effective for anomaly detection?, 2020.

[47] Hyunyoung Kim, Won-Ho Kim, Young-Youl Kim, and Hyun-Young Park. Air pollution and central nervous system disease: A review of the impact of fine particulate matter on neurological disorders. *Frontiers in Public Health*, 8, 2020.

[48] Jugal Kishore, ManishKumar Goel, and Pardeep Khanna. Understanding survival analysis: Kaplan-meier estimate. *International Journal of Ayurveda Research*, 1(4):274–278, 2010.

[49] R. Kumari, Sheetanshu, M. K. Singh, R. Jha, and N.K. Singh. Anomaly detection in network traffic using k-mean clustering. *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, Mar 2016.

[50] Thuc Duy Le, Tao Hoang, Jiuyong Li, Lin Liu, Huawen Liu, and Shu Hu. A fast pc algorithm for high dimensional causal discovery with multi-core pcs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(5):1483–1495, September 2019.

[51] Lumen Learning. The work of breathing. *Lumen Learning*, 2021.

[52] Yun-Gi Lee, Pureun-Haneul Lee, Seon-Muk Choi, Min-Hyeok An, and An-Soo Jang. Effects of air pollutants on airway diseases. *International Journal of Environmental Research and Public Health*, 18(18), 2021.

[53] Shanshan Li, Peter J. Baker, Bin B. Jalaludin, Guy B. Marks, Lyn S. Denison, and Gail M. Williams. Ambient temperature and lung function in children with asthma in australia. *European Respiratory Journal*, 43(4):1059–1066, 2014.

[54] Jiangbo Liu, Xuejiao Qian, Xueyan Jie, and Ping Jiang. Effects of different inhalant allergens on lung functions in adult patients with bronchial asthma. *Immunity, Inflammation and Disease*, 11(12):e1118, 2023.

[55] Manel Lujan, Cristina Lalmolda, and Begum Ergan. Basic concepts for tidal volume and leakage estimation in non-invasive ventilation. *Turkish Thoracic Journal*, 20(2):140–146, Oct 2019.

[56] Elaine K Luo. Humidity and asthma: What's the link?, Jun 2019.

[57] Sharan Maiya. Investigating the respiratory rate response to pm2.5 exposure in asthmatic adolescents. Honours year thesis, School of Informatics, University of Edinburgh, 2020.

[58] Ioannis Manisalidis, Elisavet Stavropoulou, Agathangelos Stavropoulos, and Eugenia Bezirtzoglou. Environmental and health impacts of air pollution: A review. *Frontiers in Public Health*, 8, 2020.

[59] Howard J. Miller. Chapter 40 - chronic obstructive pulmonary disease. In James Duke, editor, *Anesthesia Secrets (Fourth Edition)*, pages 278–284. Mosby, Philadelphia, fourth edition edition, 2011.

[60] Venkata Srikanth Nallanthighal, Zohreh Mostaani, Aki Härmä, Helmer Strik, and Mathew Magimai-Doss. Deep learning architectures for estimating breathing signal and respiratory parameters from speech recordings. *Neural Networks*, 141:211–224, 2021.

[61] Chris Fook Sheng Ng, Masahiro Hashizume, Yasushi Obase, Masataka Doi, Kei Tamura, Shinya Tomari, Tetsuya Kawano, Chizu Fukushima, Hiroto Matsuse, Yeonseung Chung, Yoonhee Kim, Kenichi Kunimitsu, Shigeru Kohno, and Hiroshi Mukae. Associations of chemical composition and sources of pm2.5 with lung function of severe asthmatic adults in a low air pollution environment of urban nagasaki, japan. *Environmental Pollution*, 252:599–606, 2019.

[62] Nicholas J. Osborne, Ian Alcock, Benedict W. Wheeler, Shakoor Hajat, Christophe Sarran, Yolanda Clewlow, Rachel N. McInnes, Deborah Hemming, Mathew White, Sotiris Vardoulakis, and Lora E. Fleming. Pollen exposure and hospitalization due to asthma exacerbations: daily time series in a european city. *International Journal of Biometeorology*, 61(10):1837–1848, October 2017. Publisher Copyright: © 2017, The Author(s).

[63] Sharon B. Park and Divya Khattar. Tachypnea. *National Library of Medicine*, Feb 2023.

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[65] Bersain A. Reyes, Natasa Reljin, Youngsun Kong, Yunyoung Nam, Sangho Ha, and Ki H. Chon. Employing an incentive spirometer to calibrate tidal volumes estimated from a smartphone camera. *Sensors*, 16(3), 2016.

[66] Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 1388–1397. PMLR, 03–06 Aug 2020.

[67] Jakob Runge. Tigramite, Apr 2023.

[68] Jakob Runge, Vladimir Petoukhov, Jonathan F. Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6(1), Oct 2015.

[69] Julia Schonthaler. Exploring asthmatic patients' sensitivity to airborne particulates through causal analysis: Investigating various causes and effects. Honours year thesis, School of Informatics, University of Edinburgh, 2024.

[70] G. Shaddick, Matthew Thomas, P. Mudu, G. Ruggeri, and S. Gumy. Half the world's population are exposed to increasing air pollution. *npj Climate and Atmospheric Science*, 3:23, 06 2020.

[71] Yen-Ling Shen, Wen-Te Liu, Kang-Yun Lee, Hsiao-Chi Chuang, Hua-Wei Chen, and Kai-Jen Chuang. Association of pm2.5 with sleep-disordered breathing from a population-based study in northern taiwan urban areas. *Environmental Pollution*, 233:109–113, 2018.

[72] Bhupendra Pratap Singh, Deepak Singh, Krishan Kumar, and Vinod Kumar Jain. Study of seasonal variation of pm2.5 concentration associated with meteorological parameters at residential sites in delhi, india. *Journal of Atmospheric Chemistry*, 78(3):161–176, Apr 2021.

[73] European Respiratory Society. Your lungs and exercise. *Breathe*, 12(1):97–100, 2016.

[74] Huacheng Song. The impact of factors - size fractions of airborne particles, temperature, relative humidity and activity - on coughing episodes in asthmatic adolescents. Master's thesis, School of Informatics, University of Edinburgh, 2023.

[75] Joan B Soriano, Parkes J Kendrick, Katherine R Paulson, Vinay Gupta, Elissa M Abrams, Rufus Adesoji Adedoyin, Tara Ballav Adhikari, Shailesh M Advani, Anurag Agrawal, Elham Ahmadian, and et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: A systematic analysis for the global burden of disease study 2017. *The Lancet Respiratory Medicine*, 8(6):585–596, Jun 2020.

[76] Marta Szemraj, Dimal Luta, Szymon Lason, and Jurgen van Leeuwen. Why is air quality worse at night? *Airly*, 2020.

[77] Eric J. Tchetgen Tchetgen, Isabel R. Fulcher, and Ilya Shpitser. Auto-G-Computation of Causal Effects on a Network. *Journal of the American Statistical Association*, 116(534):833–844, April 2021.

[78] Zac Teo Zi Cheng. Classification of sleep periods and positions using the respeck monitor. Master's thesis, School of Informatics, University of Edinburgh, 2018.

[79] Philippe Thunis, Alain Clappier, Matthias Beekmann, Jean Philippe Putaud, Cornelis Cuvelier, Jessie Madrazo, and Alexander de Meij. Non-linear response of pm2.5 to changes in nox and nh3 emissions in the po basin (italy): consequences for air quality plans. *Atmospheric Chemistry and Physics*, 21(12):9309–9327, Jun 2021.

[80] Angelica I. Tiotiu, Plamena Novakova, Denislava Nedeva, Herberto Jose Chong-Neto, Silviya Novakova, Paschalis Steiropoulos, and Krzysztof Kowal. Impact of air pollution on asthma outcomes. *International Journal of Environmental Research and Public Health*, 17(17), 2020.

[81] Huan Minh Tran, Feng-Jen Tsai, Yueh-Lun Lee, Jer-Hwa Chang, Li-Te Chang, Ta-Yuan Chang, Kian Fan Chung, Han-Pin Kuo, Kang-Yun Lee, Kai-Jen Chuang, and Hsiao-Chi Chuang. The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence. *Science of The Total Environment*, 898:166340, 2023.

[82] Lloyd N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Review*, 56(3):385–458, 2014.

[83] Laís Silva Vidotto, Celso Ricardo Carvalho, Alex Harvey, and Mandy Jones. Dysfunctional breathing: What do we know? *Jornal Brasileiro de Pneumologia*, 45(1), 2019.

[84] Hetong Wang, Laiyu Song, Wenhui Ju, Xuguang Wang, Lu Dong, Yining Zhang, Ping Ya, Chun Yang, and Fasheng Li. The acute airway inflammation induced by pm2.5 exposure and the treatment of essential oils in balb/c mice. *Scientific Reports*, 7(1), Mar 2017.

[85] J. Wanger, J. L. Clausen, A. Coates, O. F. Pedersen, V. Brusasco, F. Burgos, R. Casaburi, R. Crapo, P. Enright, C. P. M. van der Grinten, P. Gustafsson, J. Hankinson, R. Jensen, D. Johnson, N. MacIntyre, R. McKay, M. R. Miller, D. Navajas, R. Pellegrino, and G. Viegi. Standardisation of the measurement of lung volumes. *European Respiratory Journal*, 26(3):511–522, 2005.

[86] Kathleen Williams, Marina Hinojosa-Kurtzberg, and Sairam Parthasarathy. Control of breathing during mechanical ventilation: Who is the boss? *Respiratory Care*, 56(2):127–139, 2011.

[87] Hong Lei Wu, Wei Zhang Xiao, Xu Juan Xu, Yan Hong Gu, Feng Ying Lu, and Jia Hai Shi. Relationship of tidal volume to peak flow, breath rate, i:e and plateau time: Mock study. *The American Journal of the Medical Sciences*, 349(4):312–315, Apr 2015.

[88] Ailiang Xie. Effect of sleep on breathing - why recurrent apneas are only seen. *Journal of Thoracic Disease*, 4(2), 2011.

[89] Yang Xie, Hancheng Dai, Yanxu Zhang, Yazhen Wu, Tatsuya Hanaoka, and

Toshihiko Masui. Comparison of health and economic impacts of pm2.5 and ozone pollution in china. *Environment International*, 130:104881, 2019.

[90] Hongzuo Xu, Guansong Pang, Yijie Wang, and Yongjun Wang. Deep isolation forest for anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12591–12604, Jun 2023.

[91] Janelle Yorke, Anne-Marie Russell, Jeff Swigris, Caroline Shuldham, Carol Haigh, Nikki Rochnia, Jennifer Hoyle, and Paul W. Jones. Assessment of dyspnea in asthma: Validation of the dyspnea-12. *Journal of Asthma*, 48(6):602–608, Jun 2011.

Appendix A

Breath Visualizations

A.1 Naive Breath Capture Method

Figure A.1 shows noisy breathing signals where a single data point beyond the zero-crossing is identified as an entire exhalation or inhalation. This was a problem with the old, naive, breath capturing technique. Such situations must be accounted for by the breath capture method, to ensure exhalations and inhalations are appropriately identified and respiratory flow can be accurately approximated. The ”gold standard” method does not treat these as separate inhalations or exhalations.

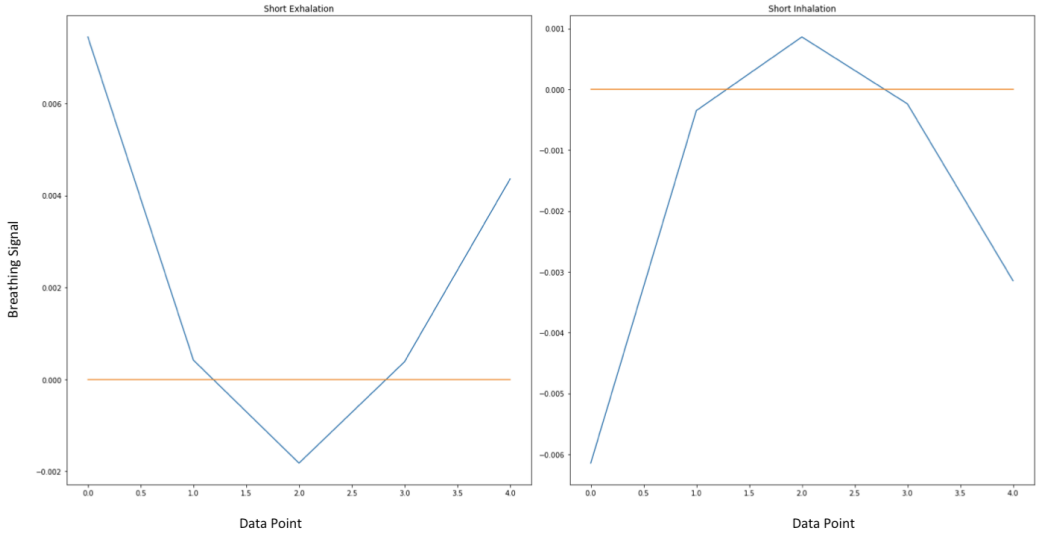


Figure A.1: Individual breathing signal data points revealing misidentified exhalation and inhalation

A.2 Gold Standard Breath Capture Method

Figure A.2 shows two breaths captured from Patient DAP134's breathing signal. In both breaths, sensor noise creates marginal zero-crossings. In the naive approach, these would result in misidentified breaths like in Appendix A.1. However, the "gold-standard" method prevents this and treats them as part of the original breath. It therefore handles sensor noise and offsets more accurately than the naive method.

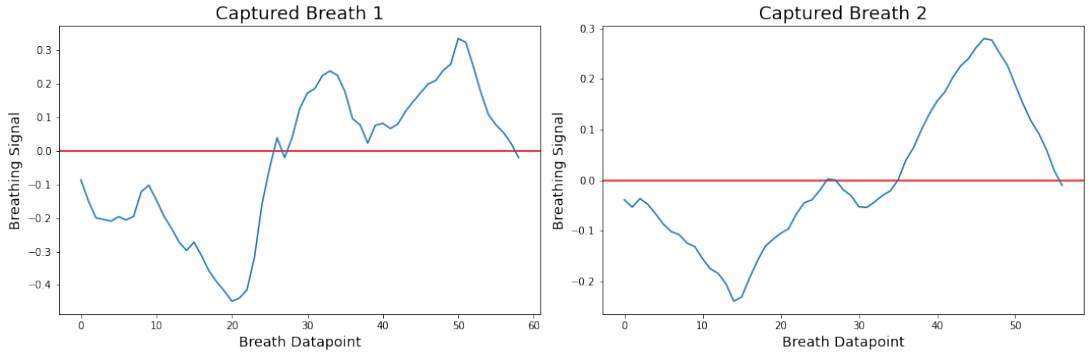


Figure A.2: Breaths Captured From Patient DAP134's Breathing Signal

Appendix B

Respiratory Flow Methodology

B.1 Rapid-Shallow Breathing

Figure B.1 shows an episode of regular respiratory flow in an asthmatic patient. The breathing signal is consistent and periodic, having uniform patterns across breaths. The breaths have consistent inspirations and expirations, indicating a stable rhythm and depth of breathing. The breaths themselves are controlled and non-rapid.

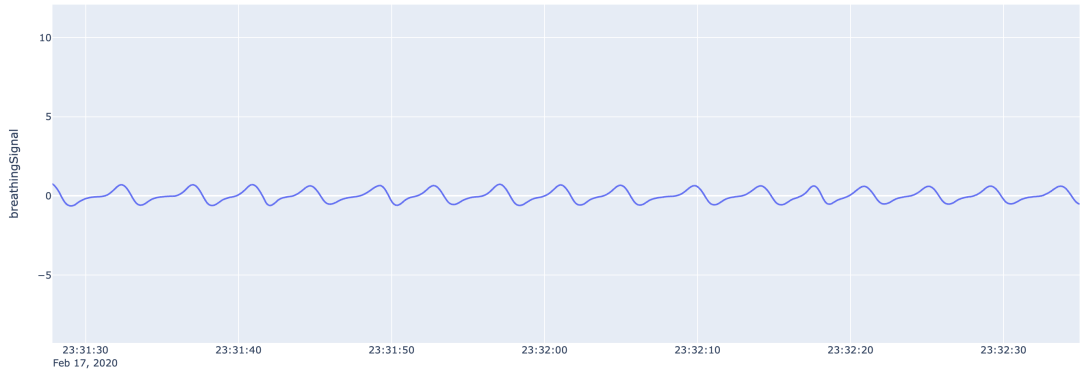


Figure B.1: Minute of Regular Respiratory Flow in Asthmatic Patient

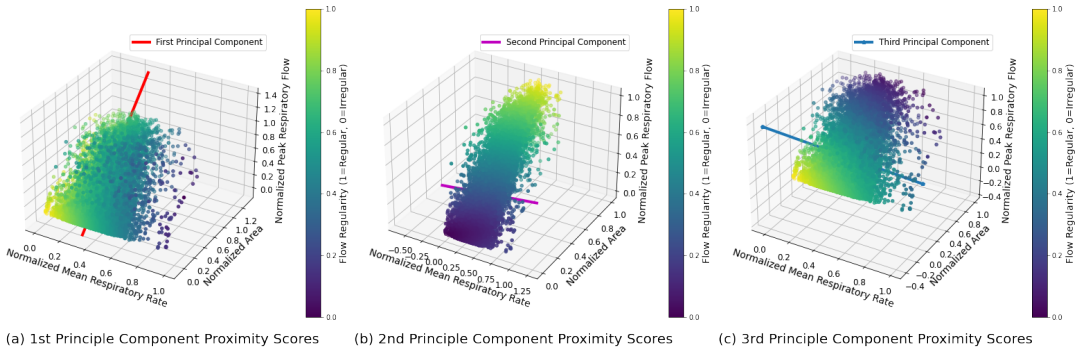
In contrast, Figure B.2 shows an episode of irregular respiratory flow in a patient. The breathing signal is very inconsistent across breaths, changing rapidly in depth and rhythm. Many breaths are rapid, often having very short and shallow (small area) inhalations or exhalations. These breaths show ”air hunger” and the state of gasping for air, commonly known as rapid-shallow breathing. Overall, the respiratory flow is very irregular and abnormal.

B.2 Distribution of Breath Features and PCA Results

This Section shows the effects of including each element in Algorithm 3 for further patients DAP063(1) in Figures B.3, B.4 and DAP061(1) in Figures B.5, B.6. In these visualizations, data points marked with a lower proximity score, represented by a darker

Figure B.2: Minute of Irregular Respiratory Flow in Asthmatic Patient

color shade, are indicative of more irregular breathing patterns. Conversely, points with a higher proximity score, denoted by a lighter color, are associated with more regular breathing patterns. The plots show that the results used to justify the elements of Algorithm 3 are generally observed across all patients. Therefore, the algorithm generalises well to all patients, producing expected results.



(a) 1st Principle Component Proximity Scores (b) 2nd Principle Component Proximity Scores (c) 3rd Principle Component Proximity Scores

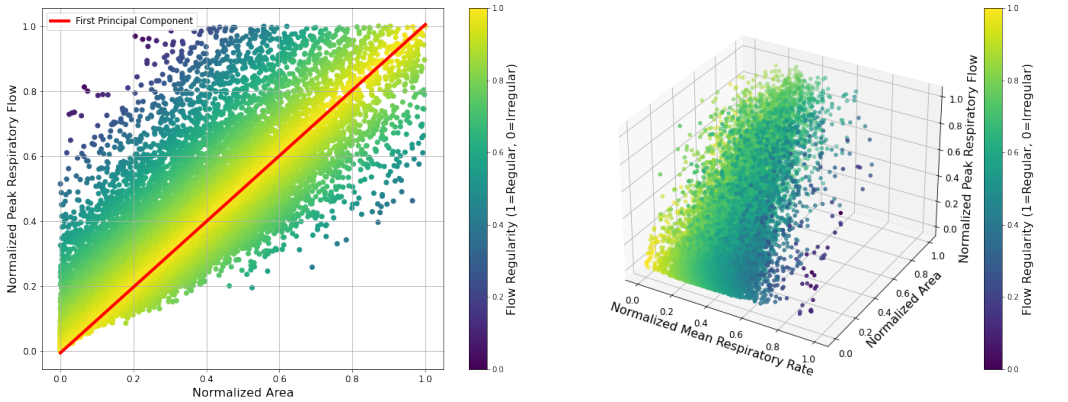
Figure B.3: Individual Metric Components For Patient DAP063(1)

B.3 Distribution of Regularity of Respiratory Flow

Figures B.7 and B.8 show the distribution of area, peak respiratory flow and respiratory rate for different bins of respiratory flow regularity. The same general trends can be observed across patients in the dataset. These are, as noted in Chapter 4, decreasing area, increasing respiratory rate and increasing peak respiratory flow as a breaths become more irregular. This conforms to the rapid-shallow expectations of irregular respiratory periods.

B.4 Mahalanobis Distance

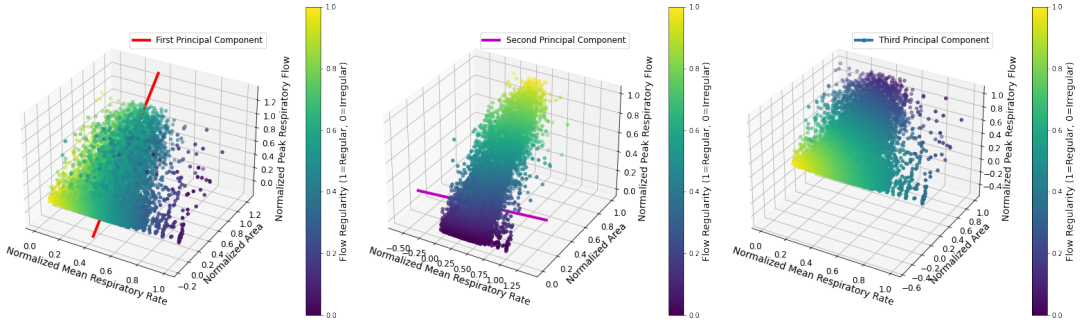
The Mahalanobis distance is a multivariate metric that measures the distance between a data-point and a distribution. It is commonly used for anomaly detection for data-points



(a) Proximity Score to 1st PC of Area and Peak Respiratory Flow

(b) Aggregated Proximity Scores to PCs of Area, Peak Respiratory Flow and Respiratory Rate

Figure B.4: Further and Aggregated Metric Components For Patient DAP063(1)

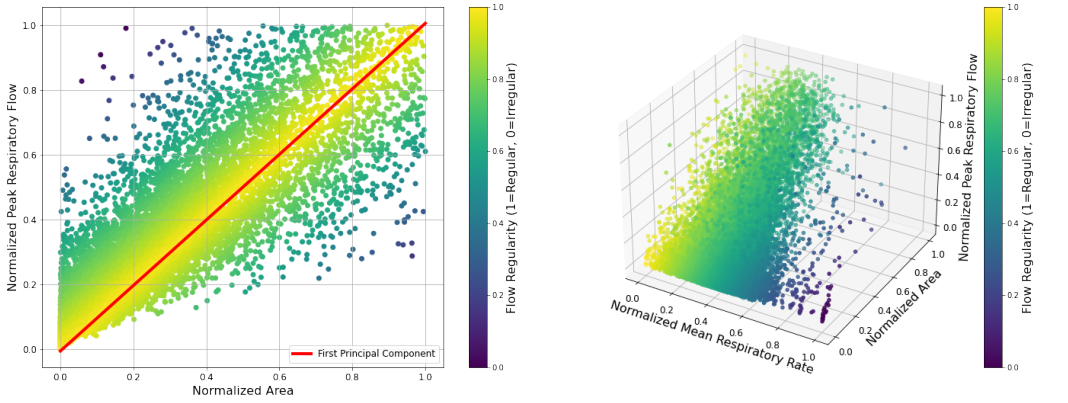


(a) 1st Principle Component Proximity Scores

(b) 2nd Principle Component Proximity Scores

(c) 3rd Principle Component Proximity Scores

Figure B.5: Individual Metric Components for patient DAP061(1)



(a) Proximity Score to 1st PC of Area and Peak Respiratory Flow

(b) Aggregated Proximity Scores to PCs of Area, Peak Respiratory Flow and Respiratory Rate

Figure B.6: Further and Aggregated Metric Components for patient DAP061(1)

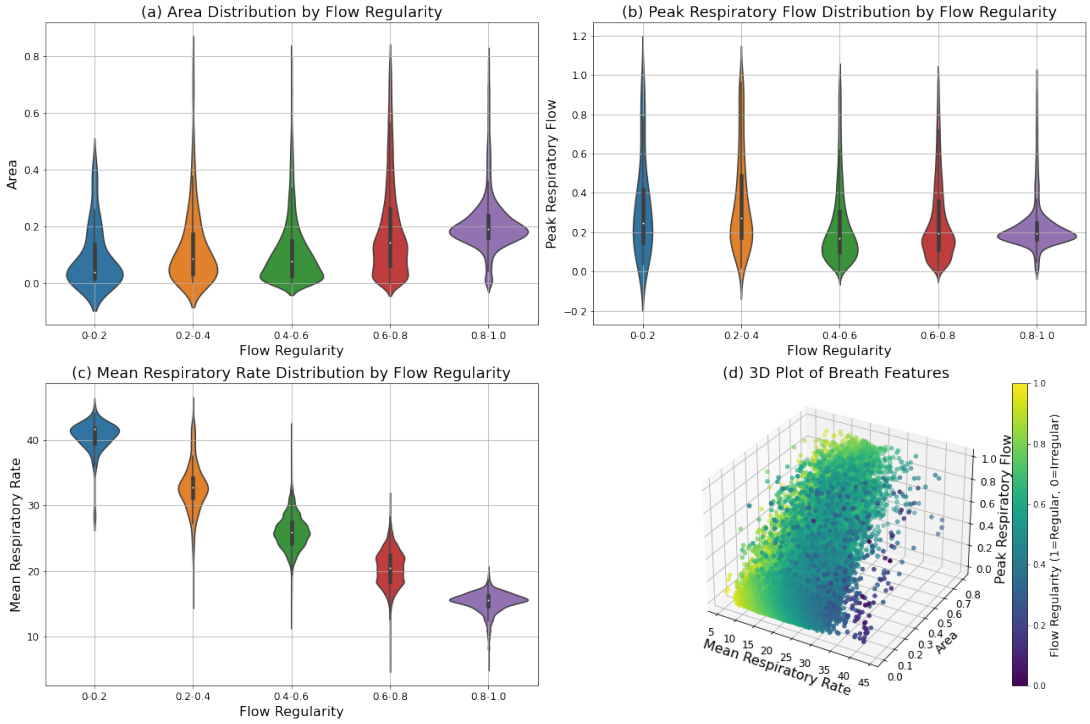


Figure B.7: Distribution of breath features across different respiratory flow regularity bins for patient DAP007(2)

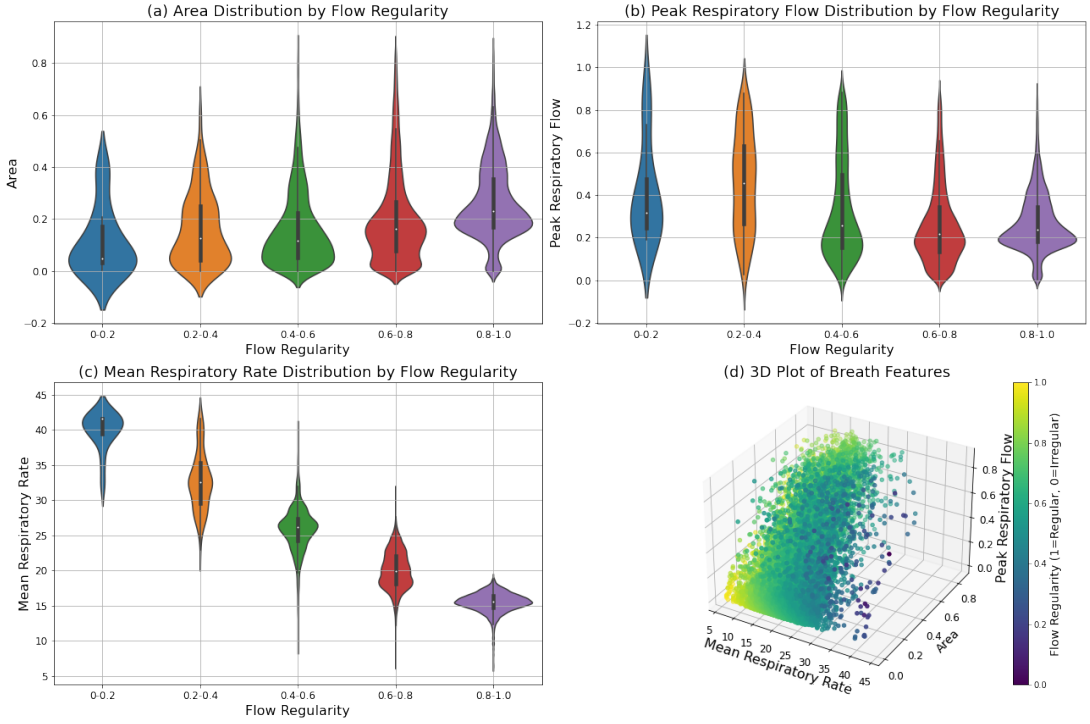


Figure B.8: Distribution of breath features across different respiratory flow regularity bins for patient DAP063(1)

that are outliers to the general distribution of the dataset. It is more accurate than Euclidean distance between the data-point and the centroid of the distribution because it takes into account the correlations of the dataset's variables. In the case of correlated dataset variables, the Euclidian distance yields misleading results. With this in mind, the Mahalanobis Distance is defined in Equation B.1.

$$D_M(\vec{p},\vec{\mu};\mathbf{Q}) = \sqrt{(\vec{p}-\vec{\mu})^T\mathbf{S}^{-1}(\vec{p}-\vec{\mu})} \tag{B.1}$$

Where:

- D_M is the Mahalanobis distance
- $\mathbf{Q}$ is the distribution of the data-points
- $\vec{p}$ is the vector point for which the distance is measured
- $\vec{\mu}$ is the mean vector of the distribution Q
- $\mathbf{S}$ is the covariance matrix of the distribution

B.5 Respiratory Flow and Activities

To better understand the regularity of respiratory flow, an investigation was done into its trends across activity types. Importantly, these activities were stationary, as the breathing signal is unreliable for high intensity activities. The findings showed that the activity type itself did not impact the regularity of respiratory flow in a consistent manner across patients. Figures B.9 and B.10 show the mean regularity of respiratory flow during different activities for Patients DAP007(2) and DAP083(1) respectively.

For many patients, all activity types had similar mean regularity of respiratory flow. However, many patients observed deviations for certain activity types that led to predominantly more regular respiratory patterns. For example, Patient DAP007 had a more regular mean respiratory flow regularity when lying down on their back and on their right side. Meanwhile, Patient DAP083(2) respired marginally more regularly on average when lying down on their left side.

That being said, these trends were not consistent across patients. Certain activities were associated with the most regular respiratory flow for some patients and then the most irregular respiratory flow for others. This lack of clear trend is however expected. Not only are all of these activities stationary and of low intensity, but also every person breathes differently. It therefore makes sense that certain positions might be marginally more comfortable for some patients but more uncomfortable for others.

Similar visualisations were made to investigate how the regularity of respiratory flow varies with time after a patient changes their activity. This is exemplified in Figure B.11, which shows how the mean respiratory flow regularity changes when Patient DAP098(2) begins to lie on their stomach. The mean respiratory flow regularity is initially consistent before marginally becoming more irregular.

However, again the findings were not consistent across patients. While many patients did not show significant deviations in their respiratory flow regularity after beginning

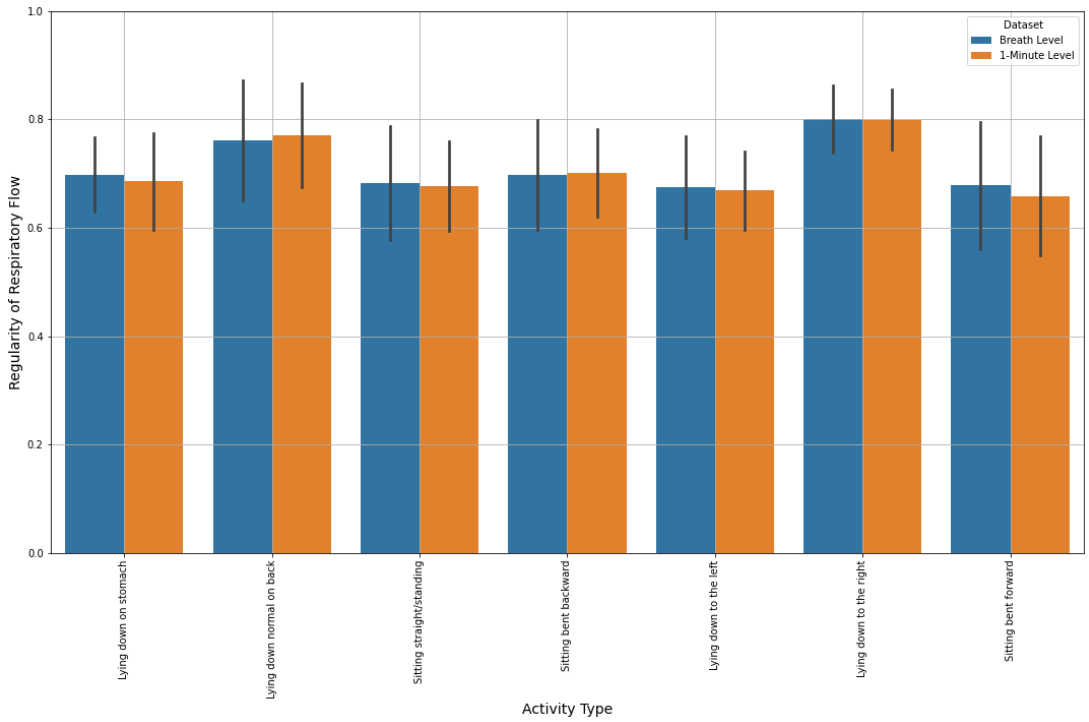


Figure B.9: Mean Regularity of Respiratory Flow During Different Activity Types For Patient DAP007(2)

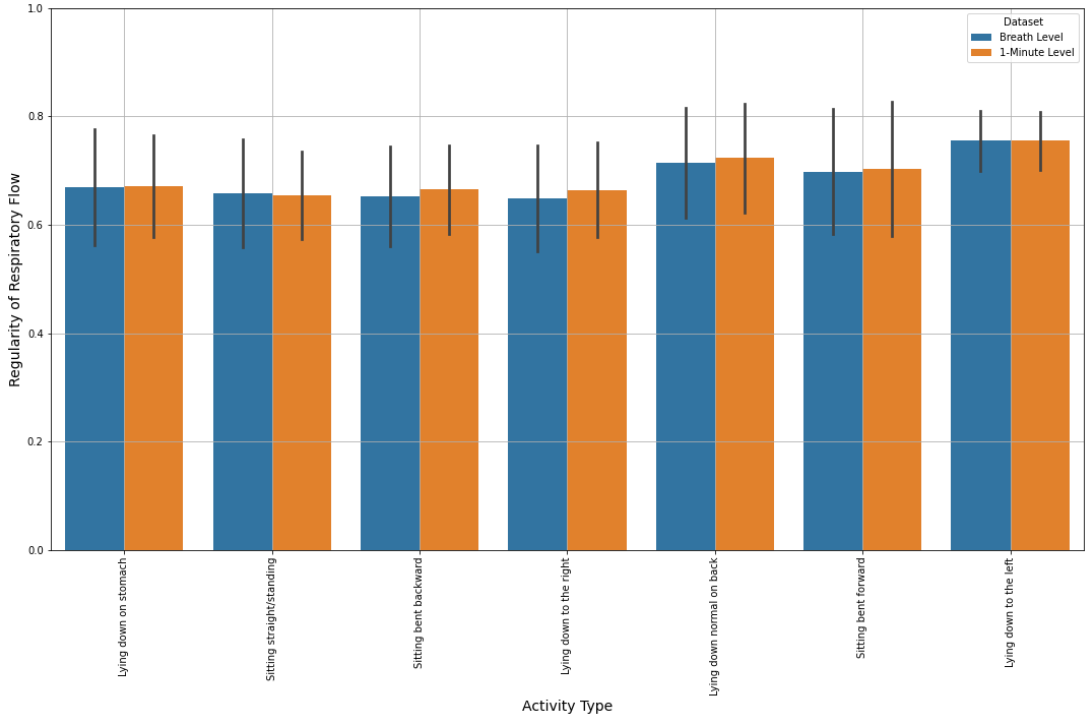


Figure B.10: Mean Regularity of Respiratory Flow During Different Activity Types For Patient DAP083(1)

certain new activities, others began to breathe substantially more regularly or irregularly. This once again highlights the differences in how every patient breathes.

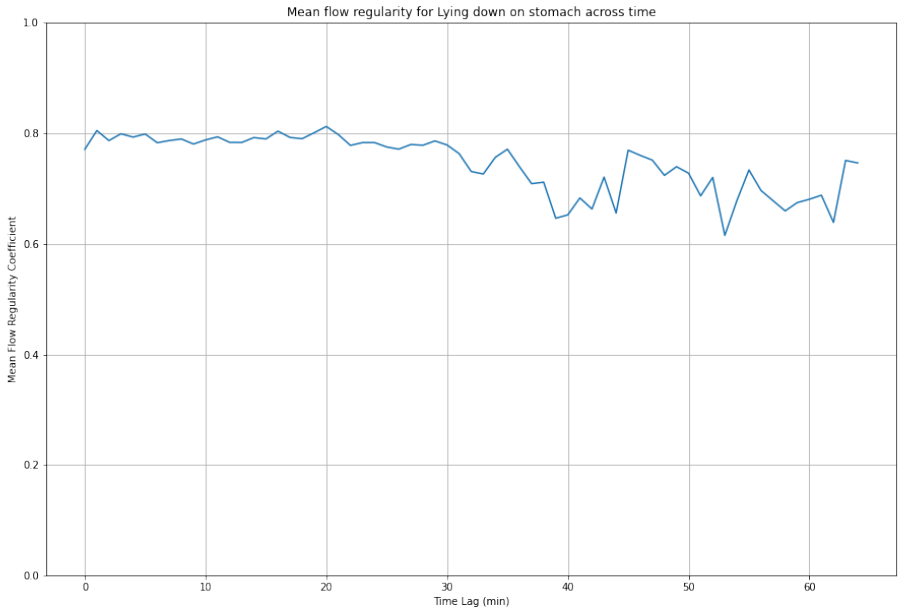


Figure B.11: Mean Regularity of Respiratory Flow When Lying On Stomach For Patient DAP098(2)

This lack of consistent trends in respiratory flow across activities was the reason this investigation was not included in the main results sections of this dissertation. Nonetheless, it may be helpful in understanding the metric and how it works.

Appendix C

Causal Discovery

C.1 PCMCI+ Setup

The PCMCI+ algorithm uses a causal graph that must be initialized with the possible causal links to test for. The setup in Figure C.1 is used for this. The edges show the potential causal relations to test for, labelled by the time lag of each adjacency. A time lag of zero is a contemporaneous link and those greater than zero are lagged links.

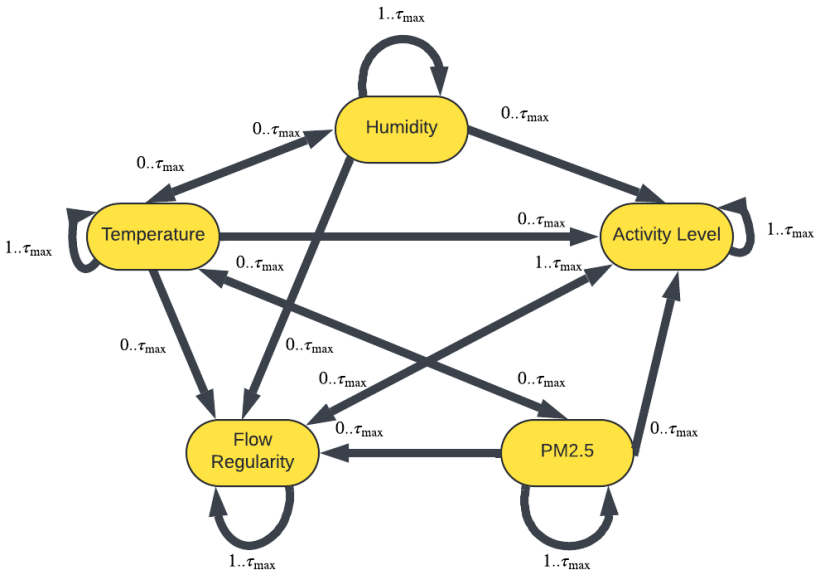


Figure C.1: PCMCI+ Initial Link Assumptions For Causal Graph Setup in Algorithm 4

C.2 Further PCMCI+ Result Analysis

C.2.1 Environmental Stimuli Causal Link Correlation Analysis

Figure C.2 shows the number of patients with each combination of causal links from the environmental stimuli to respiratory flow. Importantly, it treats the causal links as a binary variable, where a patient either has causal links or does not. Both temperature and humidity have a lot of causal effects on $PM_{2.5}$ in the PCMCI+ results of Chapter 5. The two factors are also highly related. It could therefore be hypothesized that patients that have a respiratory flow response to changes in temperature or humidity are more inclined to have a respiratory flow response to $PM_{2.5}$ as well.

However, Figure C.2 shows that the distribution of patients across the combinations of causal links from the environmental stimuli is relatively uniform with no particular trends. In the 1-hour study, exactly half of the patients with causal links from humidity to respiratory flow have causal links from $PM_{2.5}$ to respiratory flow whereas the other half does not. Similar observations can be made about patients that have causal links from temperature to respiratory flow, and also patients with no causal links to humidity and/or temperature. Therefore, these results show that there is no significant correlation between patients having respiratory flow response to $PM_{2.5}$ exposure, and patients having respiratory flow reactions to temperature or humidity.

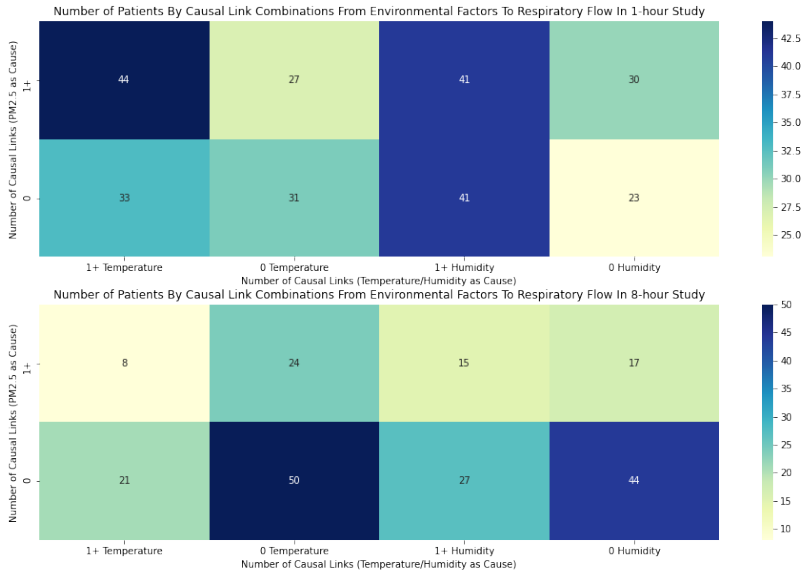


Figure C.2: Number of Patients By Causal Link Combinations from Environmental Factors To Respiratory Flow in 1-hour Study (Top) and 8-hour Study (Bottom)

C.2.2 Causal Links Across Time Lags

Figure C.3 shows the distribution of causal links from activity level to respiratory flow across time lags, for both studies. It shows that the causal links are distributed relatively

evenly across the time lags in both studies. There is a peak of 8 causal links at time lag 52. Interestingly, $PM_{2.5}$ also has a peak in causal links to respiratory flow at that time lag. That being said, there are no clear correlations between causal links from $PM_{2.5}$ and activity level to respiratory flow. An interesting observation can also be made in the 8-hour study. It seems as though the number of causal links is higher for the later lag lengths. Between 5-8 hours there are generally more links from activity level to respiratory flow than during the first three hours. This suggests a higher prevalence of delayed causal responses that a change in position or activity intensity induces on respiration in asthmatic patients.

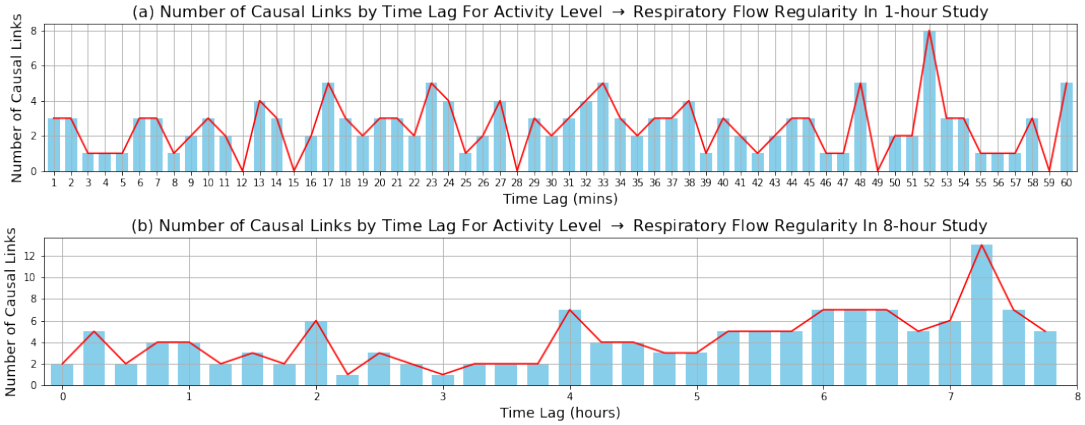


Figure C.3: Number of Causal Links Activity Level → Respiratory Flow By Time Lag In 1-hour (top) and 8-hour (bottom) Studies

Figure C.4 shows the distribution of causal links from temperature to respiratory flow across time lags, for both studies. Again, the causal links are distributed quite uniformly across the time lags with no consistent patterns across studies. There are significantly more causal links in the 1-hour study than in the 8-hour study. In fact, several time lags in the 8-hour study have no links at all. This suggests that the causal impact of temperature on respiratory flow is much more short term than that of $PM_{2.5}$.

Figure C.5 shows the distribution of causal links from humidity to respiratory flow across time lags, for both studies. Again, the causal links are distributed quite uniformly across the time lags with no consistent patterns across studies. There are also again significantly more causal links in the 1-hour study than in the 8-hour study.

C.2.3 Causal Link Significance By Cause

C.2.3.1 Temperature Results

Figure C.6 shows the statistical significance of causal links from temperature to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution. Temperature is shown to have several significant causal links to the regularity of respiratory flow for many patients. There are generally more causal links from temperature to respiratory flow than from $PM_{2.5}$ to respiratory flow, however they are also scattered across patients and lags. Overall, just over half the patients have at least one causal link from temperature to respiratory flow at a p-value threshold of 0.02.

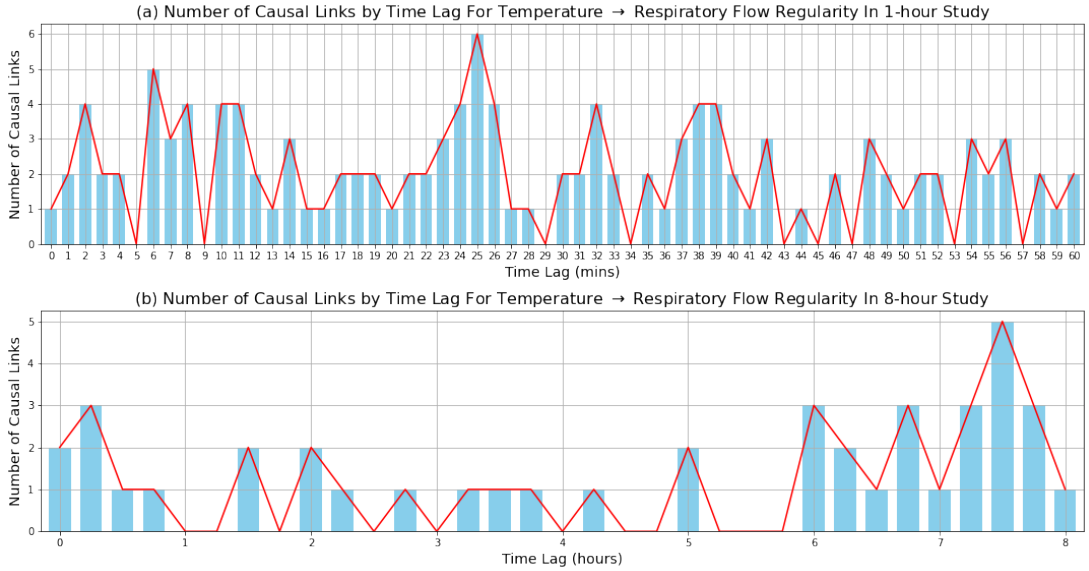


Figure C.4: Number of Causal Links Temperature → Respiratory Flow By Time Lag In 1-hour (top) and 8-hour (bottom) Studies

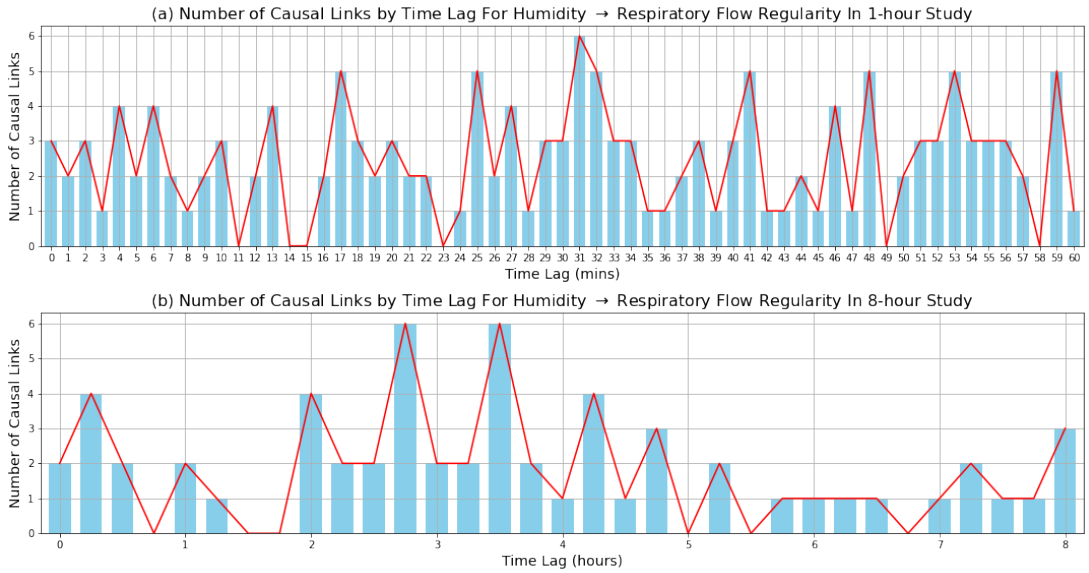


Figure C.5: Number of Causal Links Humidity → Respiratory Flow By Time Lag In 1-hour (top) and 8-hour (bottom) Studies

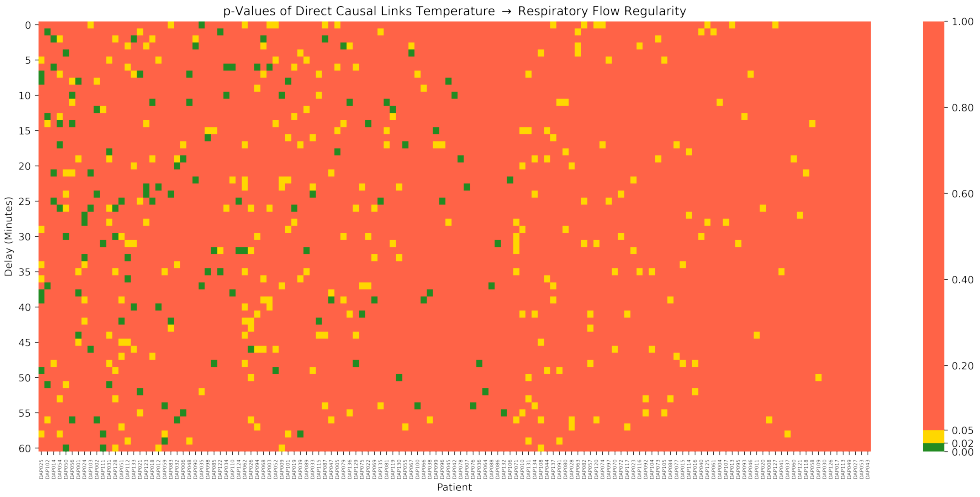


Figure C.6: Causal Links of Temperature → Respiratory Flow Over 1-Hour Period

Figure C.7 shows the statistical significance of causal links from temperature to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution. Temperature is shown to have several significant causal links to the regularity of respiratory flow for many patients. Similar to the results of $PM_{2.5}$ and respiratory flow, there are significantly less causal links from temperature to respiratory flow in the 8-hour study at 15-minute resolution. This suggests the changes in respiration are much more fine-grained and/or short term. They are therefore identified more frequently at 1-minute resolution in minute-level lags in the first hour than at 15-minute resolution in the 8-hours.

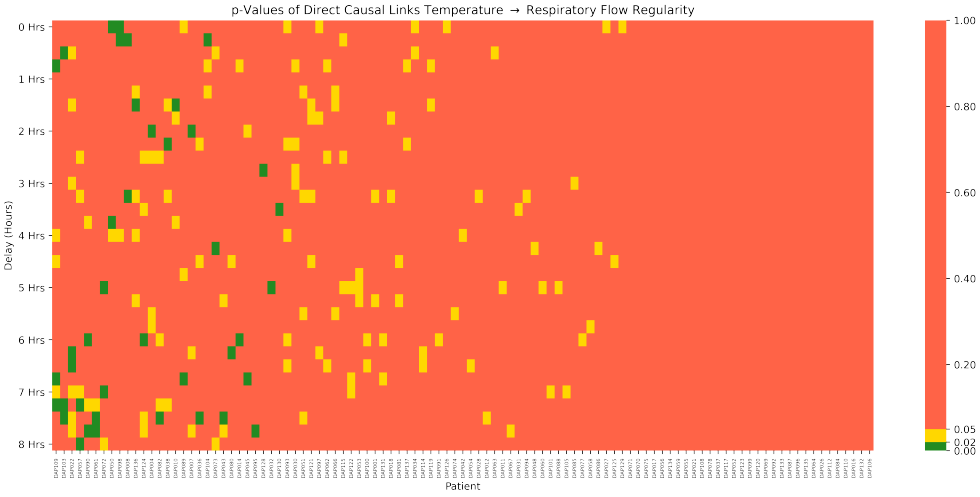


Figure C.7: Causal Links of Temperature → Respiratory Flow Over 8-Hour Period

C.2.3.2 Humidity Results

Figure C.8 shows the statistical significance of causal links from humidity to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution. Humidity is shown to have several significant causal links to the regularity of respiratory flow for many patients.

Again, the causal links are scattered across lags and patients. There appear to be marginally more patients with links from humidity to respiratory flow than from $PM_{2.5}$ to respiratory flow or temperature to respiratory flow. A majority of patients once again have causal links, showing that humidity induces causal responses in respiratory flow for many asthmatics.

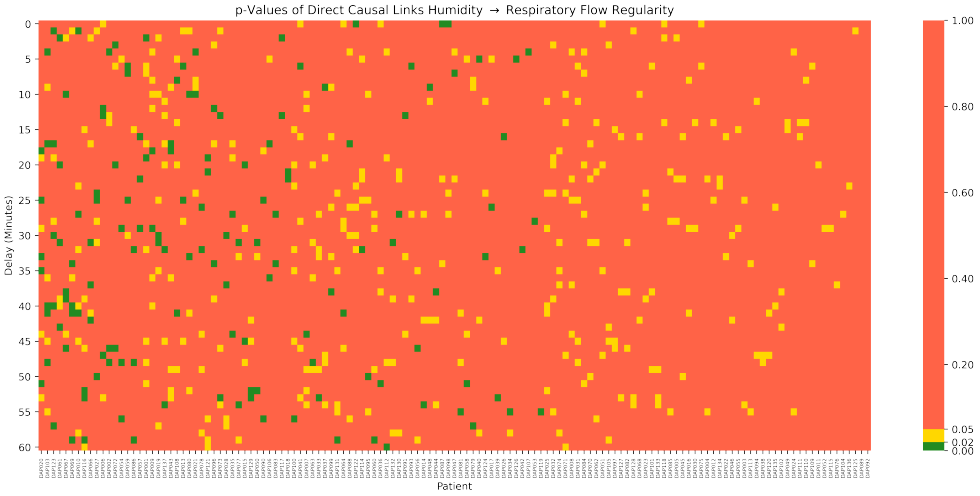


Figure C.8: Causal Links of Humidity $\rightarrow$ Respiratory Flow Over 1-Hour Period

Figure C.9 shows the statistical significance of causal links from humidity to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution. Humidity is shown to have several significant causal links to the regularity of respiratory flow for many patients. Although less patients have causal links in the 8-hour study than in the 1-hour study (similar to the other factors), there are generally substantially more links and patients with links from humidity to respiratory flow, than from temperature or $PM_{2.5}$ to respiratory flow, in the 8-hour study. This shows that the effects of humidity on respiration may be longer lasting or less fine-grained than those of $PM_{2.5}$ or temperature, meaning they are picked up in the 15-minute respiratory windows.

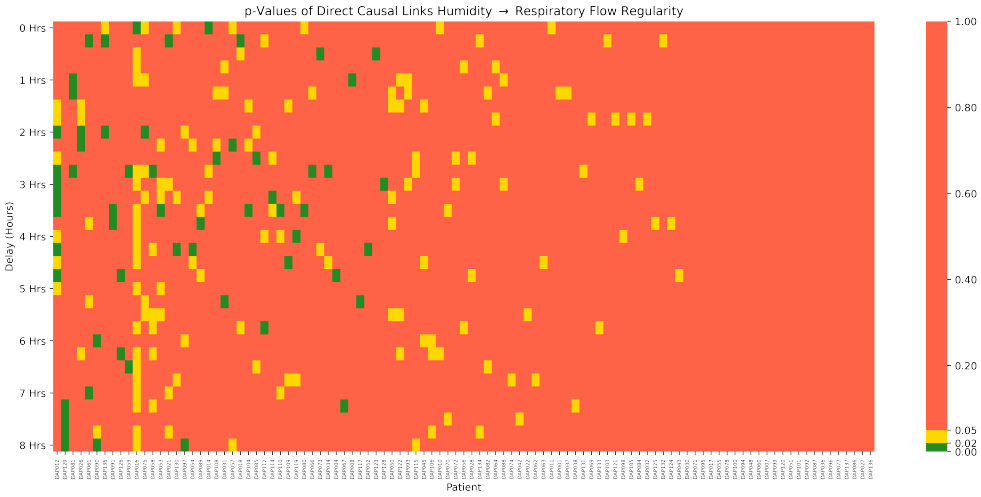


Figure C.9: Causal Links of Humidity → Respiratory Flow Over 8-Hour Period

C.2.3.3 Activity Results

Figure C.10 shows the statistical significance of causal links from activity level to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution. Activity level is shown to have several significant causal links to the regularity of respiratory flow for many patients. While the causal links seem to be scattered across patients and lags, it is observed that all patients have a causal link at time lag 0. This is because of the PCMCI+ setup, where a patient must have a contemporaneous link between activity level and respiratory flow. Doing this in the setup helps ensure the results are valid and match existing literature on the impact of physical activity on respiration. Nonetheless, a majority of patients still have lagged links from activity level to respiratory flow. This is interesting given the activities here are static (lying down, sitting and standing), showing that even variations in these activities can induce changes in respiratory flow in asthmatic patients.

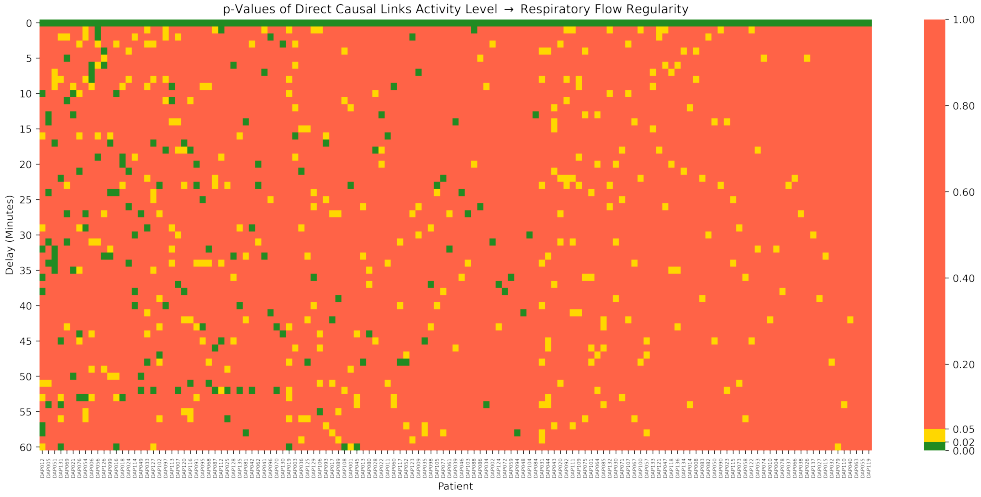


Figure C.10: Causal Links of Activity Level → Respiratory Flow Over 1-Hour Period

Figure C.11 shows the statistical significance of causal links from activity level to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution. Activity level is shown to have several significant causal links to the regularity of respiratory flow for many patients. Again, all patients have a contemporaneous link from activity level to respiratory flow by the setup of PCMCI+ to conform to existing literature. What is interesting however is that there are substantially more causal links in the later time lags. This is supported in Figure C.3. It can also be observed that there are more causal links, and patients with causal links, between activity level and respiratory flow than $PM_{2.5}$ and respiratory flow, in this 8-hour study. This shows that activity level induces effects more frequently than $PM_{2.5}$ at 15-minute resolution. However, this could also be influenced by the lower variability of $PM_{2.5}$ during the day, while activity level (e.g. the physical position we are in) changes very frequently throughout the day.

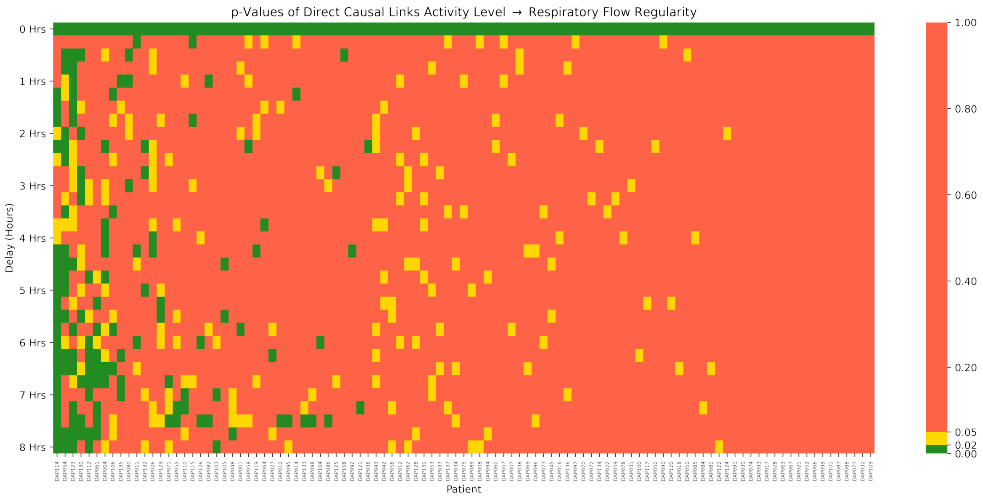


Figure C.11: Causal Links of Activity Level → Respiratory Flow Over 8-Hour Period

C.2.3.4 Respiratory Flow Results

Figure C.12 shows the statistical significance of causal links from previous respiratory flow to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution. Previous respiratory flow is shown to have several significant causal links to the regularity of respiratory flow for many patients. The setup of PCMCI+ enforces a causal link from respiratory flow to respiratory flow at lag of 1. This ensures accurate results and accurate control when determining the causal links for other factors given how we breathe currently will impact how we breathe immediately after. Nonetheless, it is observed that there are more links at the lower time lags. This further evidences the near exponential decay in number of causal links from respiratory flow to respiratory flow that is observed in both studies.

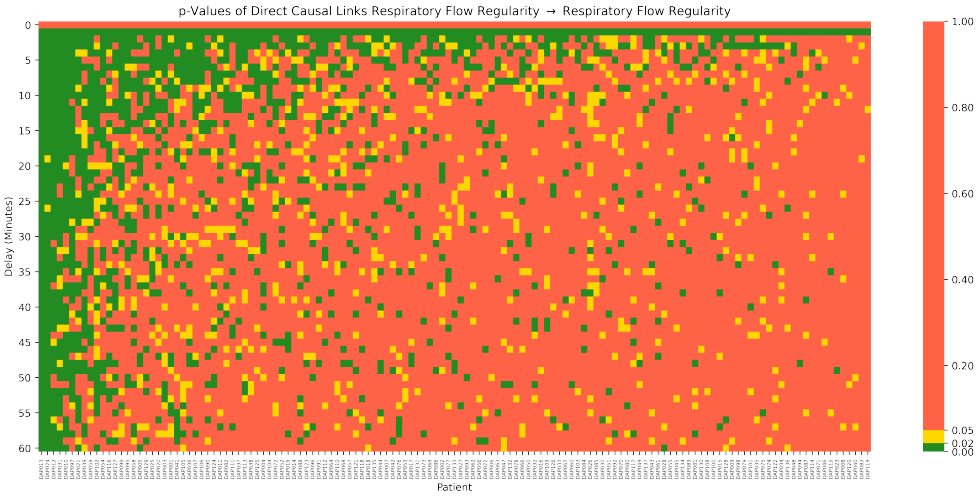


Figure C.12: Causal Links of Respiratory Flow → Respiratory Flow Over 1-Hour Period

Figure C.13 shows the statistical significance of causal links from previous respiratory flow to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution. Previous respiratory flow is shown to have several significant causal links to the regularity of respiratory flow for many patients. Again, there is a near exponential decay of number of causal links as time lag increases. This further shows that how we breathe has a decreasingly prevalent impact on how we breathe as time goes on. Intuitively, this makes sense.

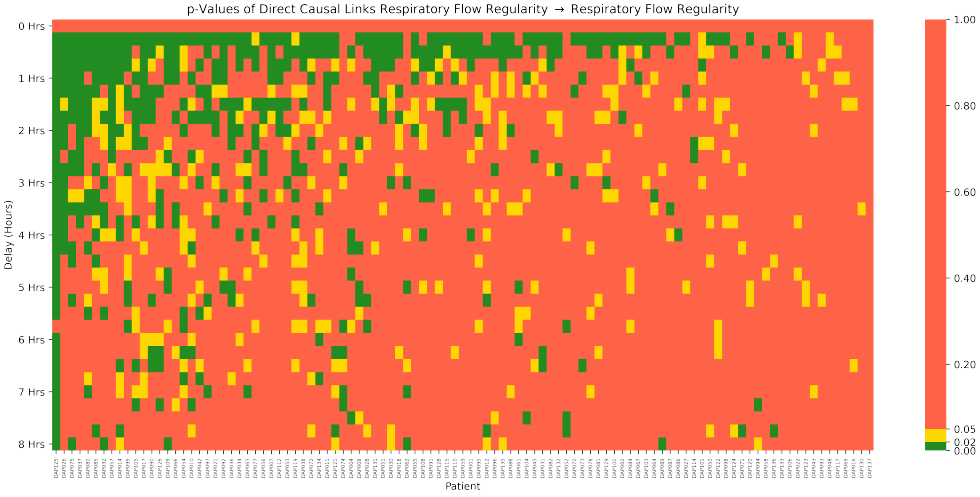


Figure C.13: Causal Links of Respiratory Flow → Respiratory Flow Over 8-Hour Period

C.3 Remaining PCMCI+ Results With Sleep Parameter

C.3.1 PM2.5 Results

Figure C.14 shows the statistical significance of causal links from $PM_{2.5}$ to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution, with the addition of the

sleep variable to the PCMCI+ algorithm. $PM_{2.5}$ remains a significant cause of change in the regularity of respiratory flow for many patients at this resolution and max lag. That being said, there are marginally less patients with causal links from $PM_{2.5}$ to respiratory flow than in the study without the sleeping parameter.

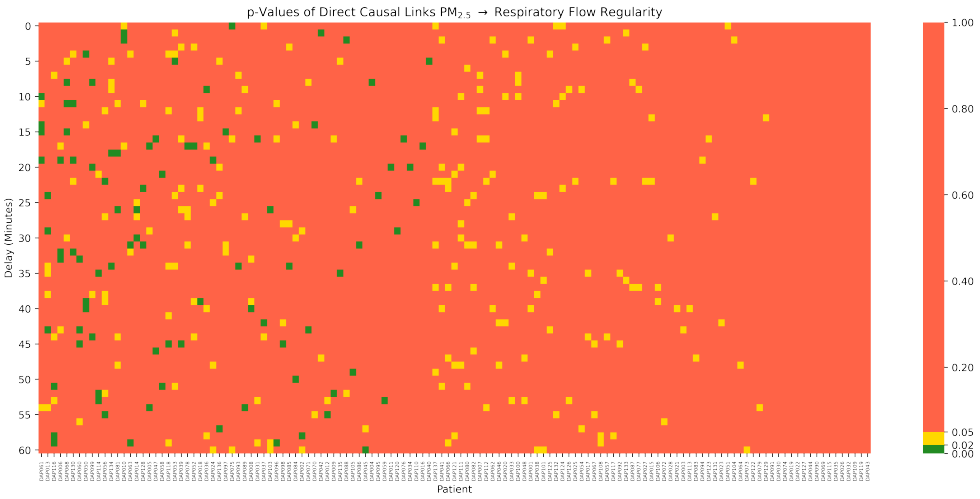


Figure C.14: Causal Links of $PM_{2.5} \rightarrow$ Respiratory Flow In 1-Hour Study With Sleep Parameter

Figure C.15 shows the statistical significance of causal links from $PM_{2.5}$ to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. $PM_{2.5}$ again remains a significant cause of change in the regularity of respiratory flow for many patients at this resolution and max lag. Interestingly, the high number of causal links of patients DAP012 and DAP071, among others, are sustained in both studies with and without the sleeping parameter.

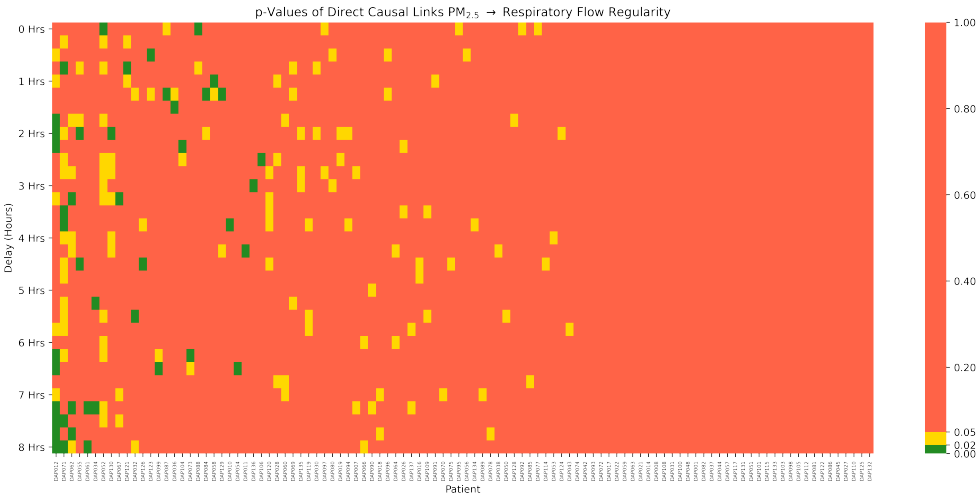


Figure C.15: Causal Links of $PM_{2.5} \rightarrow$ Respiratory Flow In 8-Hour Study With Sleep Parameter

C.3.2 Temperature Results

Figure C.16 shows the statistical significance of causal links from temperature to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Temperature is shown to be a significant cause of change in the regularity of respiratory flow for many patients. The same general trends can be seen as in the study without the sleeping parameter. One difference however is that there are less causal links and patients with causal links from temperature to respiratory flow when controlling for sleep than without. This suggests that certain causal links may have been partly explained by the patient sleeping.

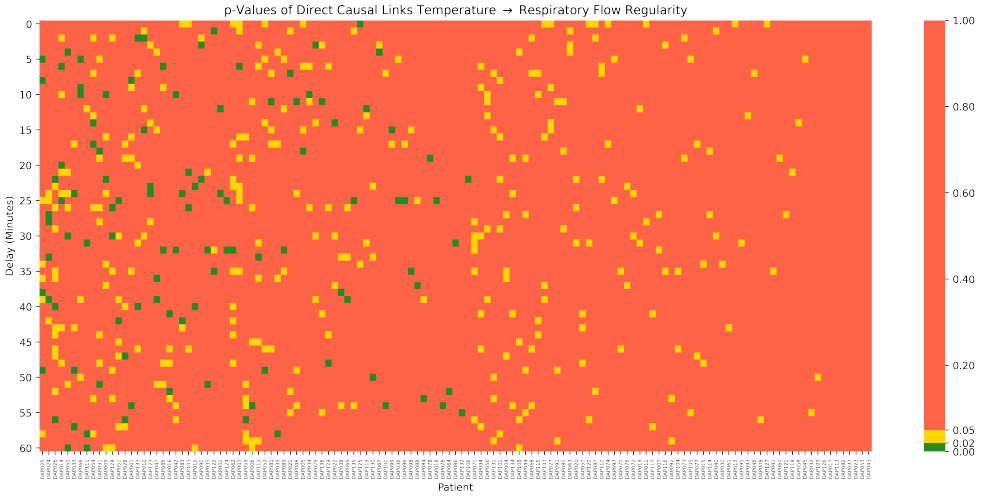


Figure C.16: Causal Links of Temperature → Respiratory Flow In 1-Hour Study With Sleep Parameter

Figure C.17 shows the statistical significance of causal links from temperature to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Temperature again is a significant cause of change in the regularity of respiratory flow for many patients. Similarly as in the other results, the general trends are the same as in the study without the sleeping parameter. However, again, there are generally less links and patients with links from temperature to respiratory flow when controlling for sleep.

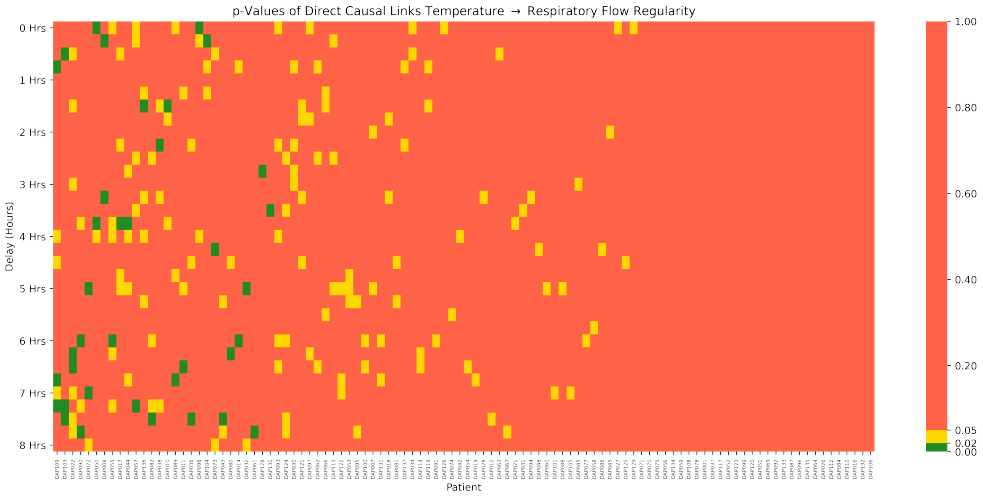


Figure C.17: Causal Links of Temperature → Respiratory Flow In 8-Hour Study With Sleep Parameter

C.3.3 Humidity Results

Figure C.18 shows the statistical significance of causal links from humidity to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Humidity is shown to be a significant cause of change in the regularity of respiratory flow for many patients. Most patients uphold their causal links from the study without the sleeping parameter in this study too.

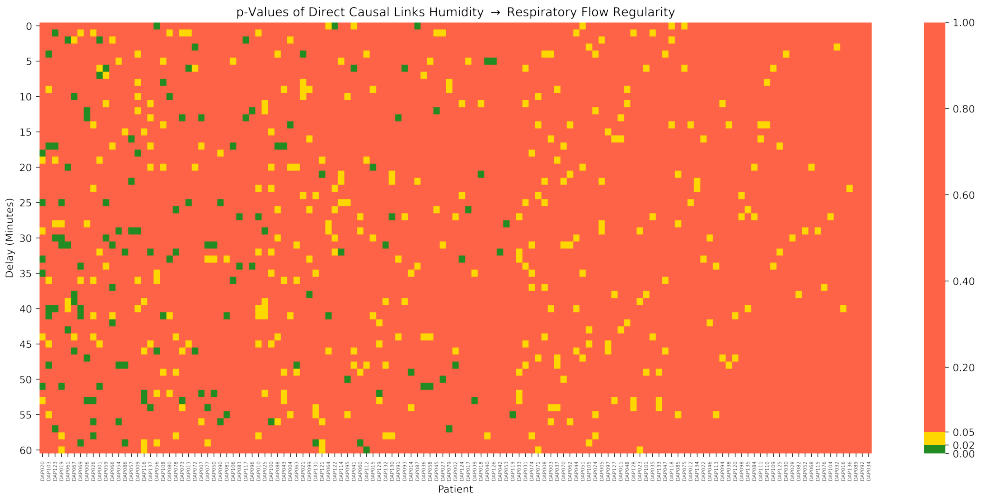


Figure C.18: Causal Links of Humidity → Respiratory Flow In 1-Hour Study With Sleep Parameter

Figure C.19 shows the statistical significance of causal links from Humidity to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Similarly as before, humidity remains a

significant cause of causal responses in respiratory flow, with many patients having the exact same causal links as in the study without the sleep parameter.

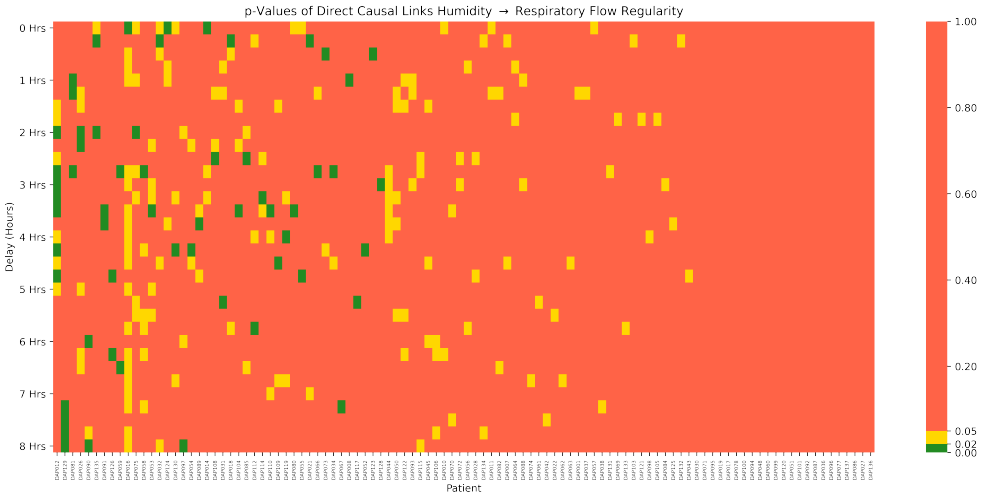


Figure C.19: Causal Links of Humidity $\rightarrow$ Respiratory Flow In 8-Hour Study With Sleep Parameter

C.3.4 Respiratory Flow Results

Figure C.20 shows the statistical significance of causal links from respiratory flow to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Meanwhile, Figure C.21 shows the statistical significance of causal links from respiratory flow to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. A similar trend is seen in these studies as in the studies without the sleep parameter. In both the 1-hour and 8-hour studies there is a near exponential decay in the number of causal links as time lag increases after exposure. Moreover, almost all patients have at least two causal links between prior respiratory flow and current respiratory flow. This again shows that breathing is not just a 1-lag relationship, but instead current breathing can impact breathing across various time lags in the future.

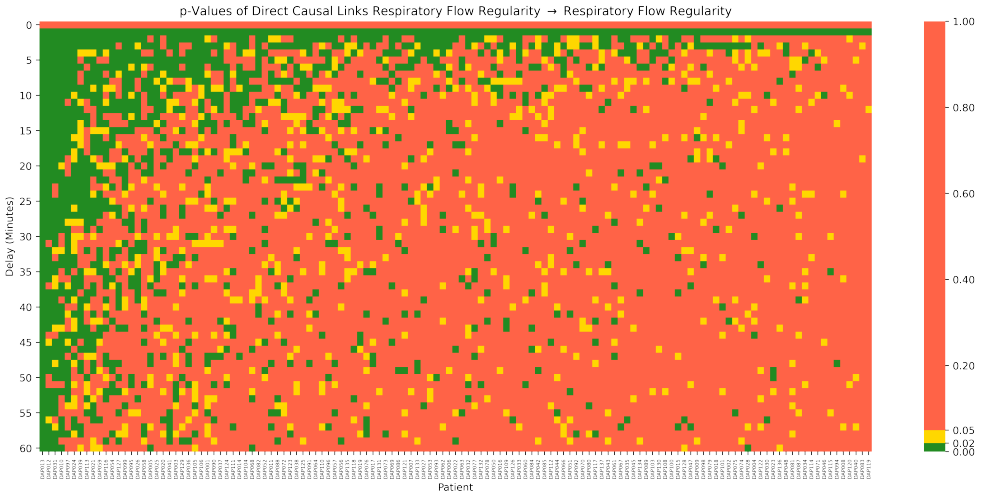


Figure C.20: Causal Links of Respiratory Flow → Respiratory Flow In 1-Hour Study With Sleep Parameter

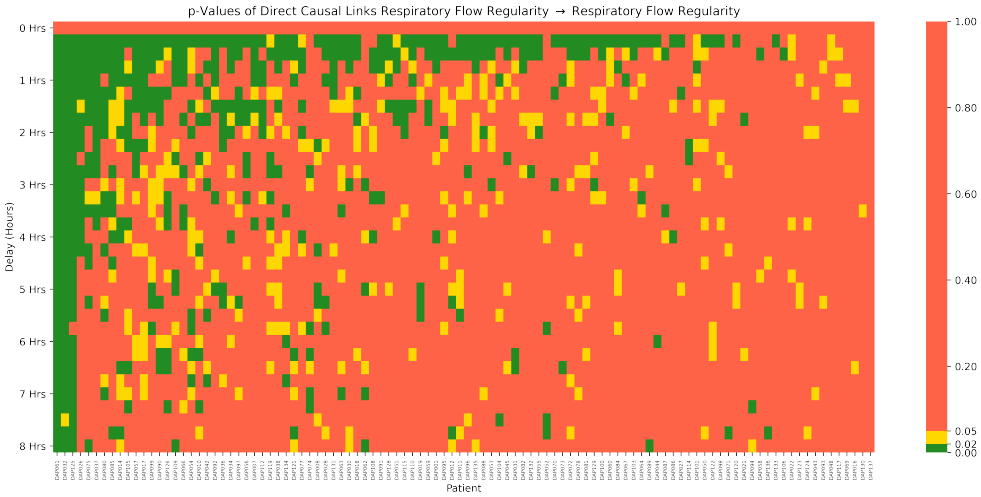


Figure C.21: Causal Links of Respiratory Flow → Respiratory Flow In 8-Hour Study With Sleep Parameter

C.3.5 Activity Level Results

Figure C.22 shows the statistical significance of causal links from activity level to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Figure C.23 shows the statistical significance of causal links from activity level to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. In both studies, there appear to be substantially more causal links from activity level to respiratory flow than from the environmental factors to respiratory flow. This is analogous to the results without controlling for sleeping. In fact, a lot of patients uphold same causal link trends in both the study with the sleep parameter and

the study without. This hints at the fact that during sleep just like during wakefulness, changing activity level by changing resting position can induce changes in the regularity of respiratory flow.

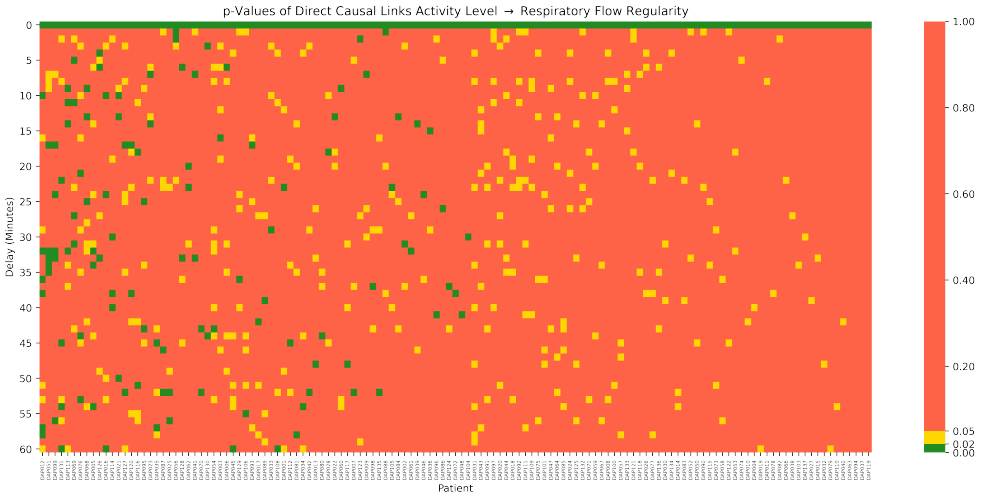


Figure C.22: Causal Links of Activity Level → Respiratory Flow In 1-Hour Study With Sleep Parameter

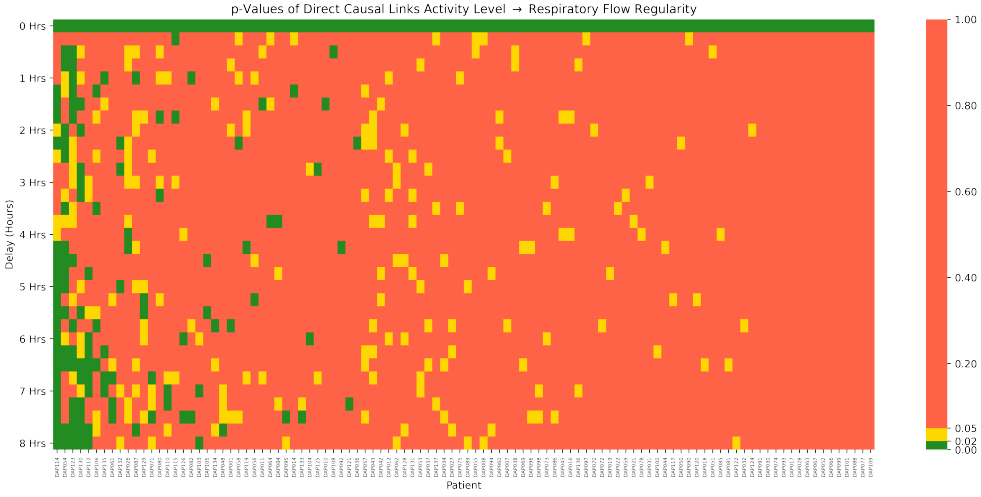


Figure C.23: Causal Links of Activity Level → Respiratory Flow In 8-Hour Study With Sleep Parameter

C.3.6 Sleep Parameter Results

Figure C.24 shows the statistical significance of causal links from the sleeping parameter to respiratory flow over 0 to 60 minute lag lengths at 1-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm. Figure C.25 shows the statistical significance of causal links from the sleeping parameter to respiratory flow over 0 to 8 hour lag lengths at 15-minute resolution, with the addition of the sleep variable to the PCMCI+ algorithm.

An interesting observation is that majority of patients in the 1-hour study have a contemporaneous link between the sleep parameter and respiratory flow. This shows that whether we are in a state of wakefulness or sleep can immediately impact our respiratory activities. Overall, a majority of patients have causal links in the 1-hour study and several patients have causal links in the 8-hour study. This shows that whether sleeping has a significant causal impact on respiratory flow and changes the respiratory dynamics in asthmatic patients.

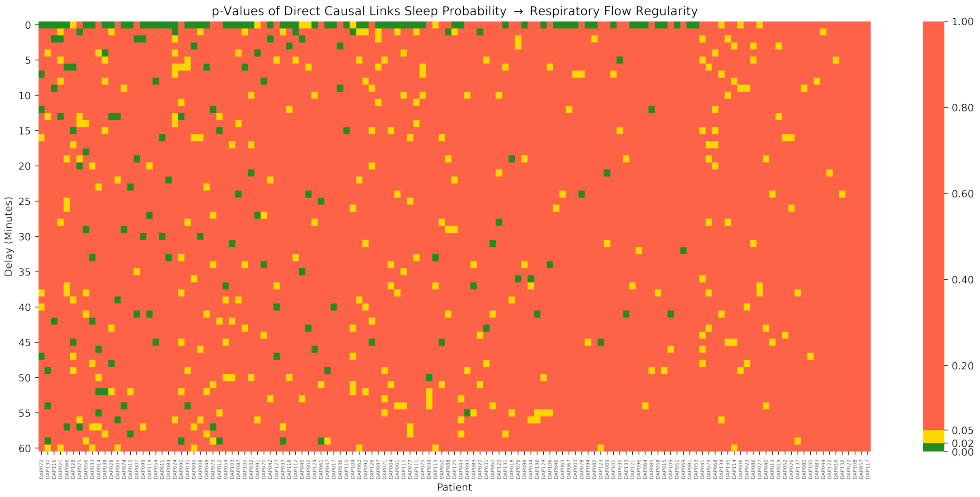


Figure C.24: Causal Links of Sleep Parameter → Respiratory Flow In 1-Hour Study

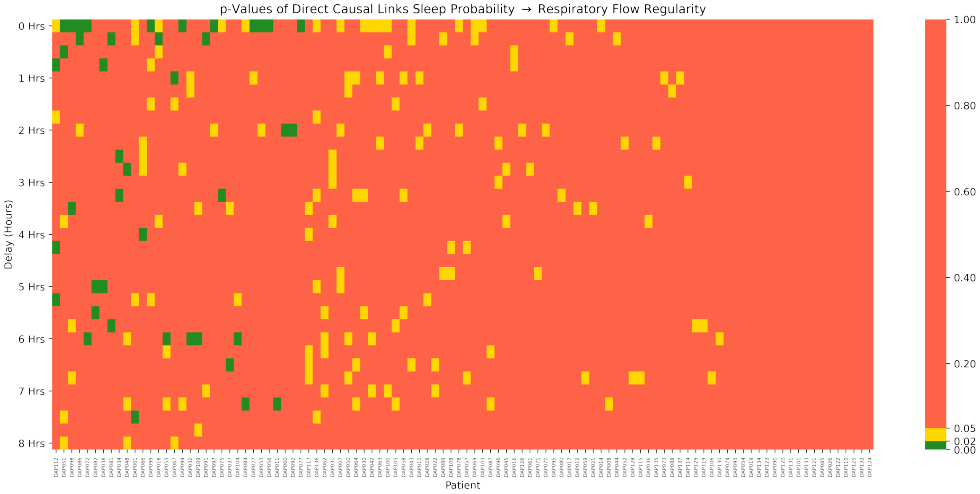


Figure C.25: Causal Links of Sleep Parameter → Respiratory Flow In 8-Hour Study

Appendix D

Further Causal Effect Results

D.1 Monthly Comparison To Respiratory Rate

Figure D.1 shows the mean and the range of the direct causal effect strengths from $PM_{2.5}$ to respiratory rate across months for the 1-hour study. Similarly, Figure D.2 shows the mean and the range of the direct causal effect strengths from $PM_{2.5}$ to respiratory rate across months for the 8-hour study. The causal effect estimation results are computed in [69] using the same methodology as in Chapter 6. The only difference is that the absolute direct effect strength was not capped at 1. This is because, unlike in the respiratory flow effect estimations, a lot of effect strengths exceeded this threshold. Not capping the magnitude therefore ensured a more complete understanding and comparison between the two respiratory outcomes. While the general trends between these plots and Figures 6.7 and 6.8 can be compared, the strengths themselves cannot. This is because the respiratory flow is a continuous measure of regularity between 0-1, while respiratory rate is typically between 10 and 25 breaths per minute. Therefore, the scales of the respiratory outcomes are not comparable, making the raw effect strengths themselves not directly comparable either.

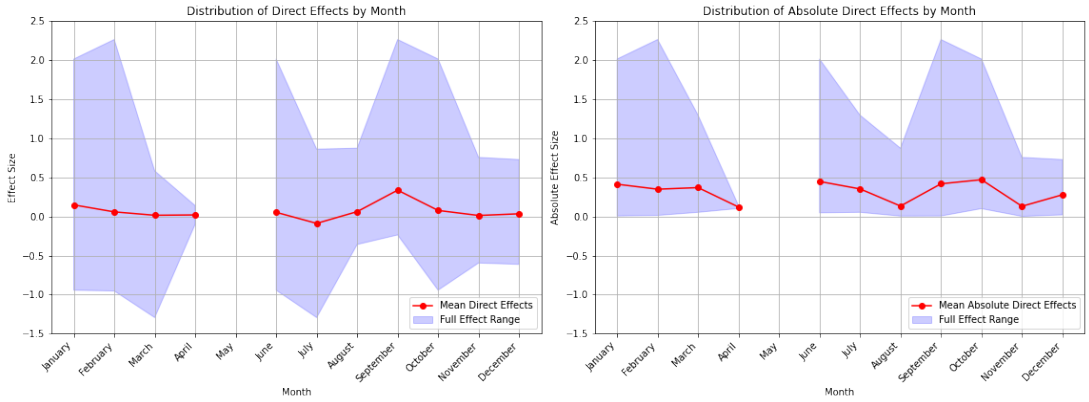


Figure D.1: Range and Mean of Direct $PM_{2.5} \rightarrow$ Respiratory Rate Causal Effect Strength Across Months in 1-hour Study. Raw Results Taken From [69]

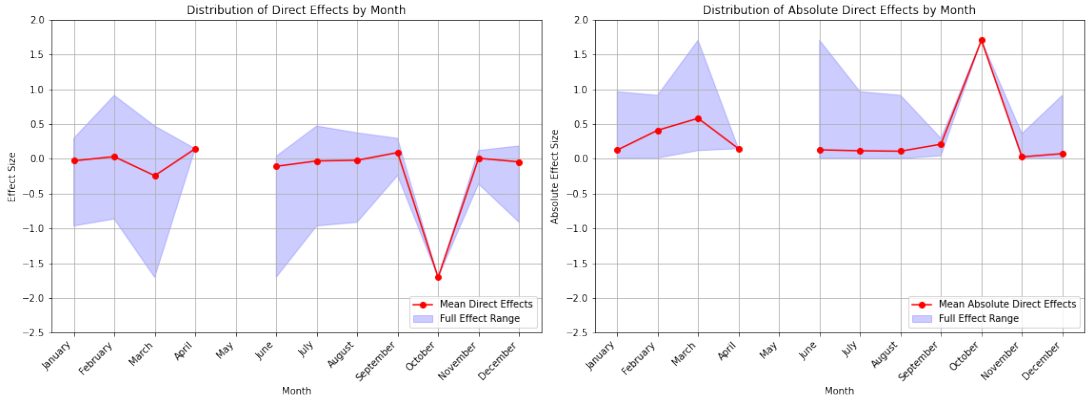


Figure D.2: Range and Mean of Direct $PM_{2.5} \rightarrow$ Respiratory Rate Causal Effect Strength Across Months in 8-hour Study. Raw Results Taken From [69]