IMPACT OF COVID-19, INFLUENZA, AND AIR POLLUTION ON MORTALITY

YANWEN LIU

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Applied Mathematics

Impact of Covid-19, Influenza, and Air Pollution on Mortality

Yanwen LIU

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

May 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Yanwen LIU (Name of student)

Abstract

Associations between various influenza strains (H3N2, H1N1, and influenza B), different air pollutants (PM2.5, PM10, ozone, etc.) and mortality have been a major concern in the past decade. The worldwide outbreak of COVID-19 in early 2020 made it urgent to explore the mortality attributed to the disease. Meanwhile, the influences of influenza and air pollution on mortality keep changing during the pandemic period and the post pandemic era. This thesis will study the impact of COVID-19, influenza, and air pollution on mortality by deploying widely adopted statistical methods, such as generalized additive models, maximum likelihood estimation and generalized linear model, along with state-of-the-art data science methods such as eXtreme Gradient Boosting and neural network.

Chapter 2 used death certificates dataset provided by Peru Ministry of Health and reported the heterogeneity of the median age of all-cause mortality and the daily pattern of cause-specific mortality directly and indirectly related to COVID in Peru. An assumption that most of the indirectly excess death in Peru were primarily caused by circulatory system diseases was raised.

In Chapter 3, exposure history and symptom onset date for patients infected by different SARS-CoV-2 variants was collected. Maximum likelihood estimators of mean and standard deviation of the distributions of incubation periods were

calculated. For the Delta variant, the mean incubation period was 6.54 days (95% CI: 5.28 – 7.68), while for the Omicron variant, it was significantly shorter at 3.43 days (95% CI: 2.47 – 3.76). Thus, a 7-day quarantine may be more effective during Omicron predominance.

Moreover, a metric called daily exceedance concentration hours (DECH) was reinvestigated in Chapter 4. Generalized additive models with quasi-Poisson distribution links were fitted to calculate relative effects of DECH levels on mortality risk across the disease groups. A fairly consistent size of the association between DECH levels and mortality risk was found with a less significance during the COVID pandemic period.

Futhermore, eXtreme Gradient Boosting was adopted in Chapter 5 to estimate the excess mortality attributable to air pollutants and influenza, during the pre-pandemic and pandemic period in Hong Kong. In the first two years of the COVID-19 pandemic, 8,762 (95% confidence interval, 7,503 – 9,993), and 12,496 (11,718 – 13,332) excess all-cause deaths were estimated. A notable shift in disease burden attributable to influenza and air pollutants was observed in the pandemic period, suggesting that mortality directly and indirectly caused by COVID-19 shall be considered when assessing the global and regional burden of the COVID-19 pandemic.

Finally, in Chapter 6, suggestions for future research on influenza forecasting were discussed. Human Influenza hemagglutinin (HA) and Neuraminidase (NA) A(H3N2) sequences and their association with influenza spread were reviewed. Deep Learning Framework was introduced to handle both sequencing and time series data.

Publications arising from the thesis

[1] Yanwen Liu et al. “All-cause mortality during the COVID-19 pandemic in Peru”. In: *IJID regions* 5 (2022), pp. 177–179.

[2] Yanwen Liu et al. “Estimating the incubation period of SARS-CoV-2 Omicron BA. 1 variant in comparison with that during the Delta variant dominance in South Korea”. In: *One Health* 15 (2022), p. 100425.

[3] Yanwen Liu et al. “Trends in the effects of ambient PM 2.5 concentration on mortality risk in Hong Kong, China [preprint]”. In: (2022).

[4] Yanwen Liu et al. “Change in disease burden associated with influenza and air pollutants during the COVID-19 pandemic in Hong Kong”. In: *Digital Health* 10 (2024), p. 20552076241261892.

Acknowledgements

I would like to extend my deepest gratitude to those who played a pivotal role in the completion of this work. First and foremost, my heartfelt thanks go to my supervisor, Prof. He, whose insightful guidance, persistent encouragement, and unwavering commitment shaped my research. I am also profoundly grateful to my co-supervisor, Dr. Li, whose innovative ideas enriched this work in countless ways. In addition, I wish to thank the research associate, Dr. Avilov, who worked diligently alongside the team, providing comments at almost every meeting.

I must also acknowledge the incredible support and collaboration of colleagues. The vibrant discussions, shared knowledge, and spirit of cooperation within our group have not only contributed to this publication but have also inspired my continuous professional growth.

On a personal note, I would not be here without the love and strength provided by my family. I am eternally thankful to my parents for their boundless encouragement, sacrifices, and belief in my potential, which have guided me through every challenge. To my wife, thank you for your unwavering support, understanding, and love.

Each one of you has left an indelible mark on my academic and personal life, and for that, I am sincerely grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 COVID-19 pandemic

The COVID-19 pandemic, initiated by the emergence of the novel coronavirus SARS-CoV-2 in late 2019, rapidly developed into a global outbreak that overwhelmed health systems and led to catastrophic loss of life [1]. By the end of March 2024, the World Health Organization (WHO) reported that the COVID-19 pandemic had caused 775 million infections, and 7 million deaths around the world [2]. [3] concluded that the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) had a half rate of influenza mutation and one-fourth rate of human immunodeficiency virus mutation. And they further revealed that the rapid mutation may cause higher infectivity, transmissibility, and lower neutralization efficacy by vaccines. [4] emphasized that although the pressure of COVID pandemic will be significantly reduced, the SARS-CoV-2 will continue to circulate and re-

main a part of our lives. Meanwhile, the Long Covid, the persistent symptoms after recovering from SARS-CoV-2 infection, caused sustained health challenges and may elevate indirectly mortality risk [5]. All the evidence argued that it is essential to understand the key parameters of different variants of SARS-CoV-2 and mortality burden related to COVID whether the pandemic ended or not.

The mortality of COVID-19 has been estimated using various methods in different regions. A study conducted in Ceará, Brazil deployed Poisson regression on a cohort with 2070 people having flu-like symptoms and tested positive to COVID-19 to estimate the mortality rate [6]. [7] based on 114 studies argued that demographic factors, such as age, gender, and behavior like smoking, as well as preexisting comorbidities including chronic health conditions could increase the risk of death. [8] ensembled multiple models to calculate the excess mortality for 74 countries and territories during the pandemic in 2020 and 2021 as the difference between observed mortality and expected mortality. A total of 18.2 million deaths was estimated as excess mortality worldwide caused by COVID-19 pandemic compared to a 5.94 million reported direct mortality. However, a further investigation on the cause of death of those excess mortality remained unclear.

Meanwhile, the global response to the COVID-19 pandemic led to a sharp reduction in economic and transportation activities. Nitrogen dioxide and particulate matter levels were reduced by roughly 60% and 31% due to the lockdown events [9]. Moreover, strict travel restrictions, lockdowns, and stay-at-home orders were implemented by governments worldwide and global human mobility was disrupted by the pandemic [10]. With the lower air pollution level, it was possible that the population had less exposure to air pollution. The association between air pollution and mortality needs to be reviewed over the COVID outbreak and it is crucial to

understand whether the association between specific-cause mortalities and $PM_{2.5}$ remain significant during the pandemic and post-pandemic period.

Besides, with the emergence of COVID-19, the seasonal influenza epidemic has been reshaped. The non-pharmacological interventions (NPI) such as face masks, social distancing, school and business closures, implemented to slow COVID-19 spread eventually limited influenza transmission and led to a significant drop in influenza cases [11]. Also, the respiratory pathogen testing increased up to a 4.6 fold between pre-pandemic period and the peak of the pandemic response [12]. This change could cause a shift in the number of positive detections of seasonal influenza and positive percentage for influenza among all respiratory specimens. Moreover, a rebound of influenza activity in the post COVID-19 pandemic period was observed and possibly caused by released non-pharmacological interventions, diminished population-level immunity, and influenza virus evolution [13]. It is essential to rebuild a long-term influenza forecasting model covering the rebound of influenza activity.

1.1.2 Association between air pollution and mortality

Similar to influenza, the association between air pollution and mortality has been noticed for decades. The well-known Great Smog of London led to about 4000 deaths directly and later analyses estimated the mortality to be around 10000 to 12000 and more than 100000 individuals suffered from respiratory diseases and other complications [14]. The air pollution started to be treated as a critical public health issue. Even though in early period, the association between air pollution and mortality was not fully measured or understood, it was suspected that adverse

health outcomes were related to air pollution [15]. Later, continuous air quality monitoring was collected, and statistical methods can be applied to time series data of air pollution.

In recent decades, the air quality has maintained or even improved due to energy policies and pollution regulations and [16] expected a 75% reduction of the mean population exposure to $PM_{2.5}$ in 2040 compared to 2015. Meanwhile, the World Health Organization (WHO) has updated Air Quality Guidelines (AQG) to a stricter threshold multiple times [17, 18, 19]. The AQG were designed to provide guidance to avoid health impacts of air pollution and included 37 of the most common air pollutants including $PM_{2.5}$, NO_2, etc. [17]. Take 24-hour mean particulate matter 10 (PM_{10}) as an example, in the 2021 updated WHO Guidelines [19], the threshold was reduced from 50 μg/m^3 to 45 μg/m^3. Yet, temporal air pollution associations with daily mortality could be found with a historically low air pollution level [15]. It remained significant to investigate the association between air pollution, even at a low level, and mortality.

Unlike the early observational approaches of air pollution, epidemiological techniques and statistical modeling have been applied to quantify the association between air pollution and mortality. Meanwhile, confounders including temperature, humidity and socioeconomic factors have been controlled in the analyses [15]. Among the 37 most common air pollutants listed in AQG, most research focused on particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), sulfur dioxide SO_2, and ozone O_3. [20] reviewed research deploying statistical methods to investigate impact of incremental increases in pollutant concentrations on mortality rates and concluded that even modest increases in air pollution levels are associated with measurable rises in daily mortality. Generalized additive model (GAM) and gen-

eralized linear model (GLM) were commonly used methods to explore the health effects of air pollution in time-series analyses and capture the complex, nonlinear relationships [21]. Temporal trends and relative humidity were smoothed as covariates [22] and mortalities were usually lagged by 0–3 days when assessing the association between air pollution and mortality. Moreover, besides using daily mean concentration or daily peak of air pollutants in the analysis, daily concentration hours of air pollutants over a certain threshold were introduced to consider both exposure intensities and durations [23].

Despite the adverse health outcomes related to air pollution, [24] examined the potential impact of ozone on influenza transmission and revealed a negative association between ambient ozone and influenza activity, particularly with a one-week lag. Possible reasons include the strong oxidizing properties of ozone and the enhancement of immune system while exposing to ambient ozone. This research indicated that the air pollution can be jointly investigated with influenza on mortality.

1.1.3 Forecasting Influenza Epidemics and Mortality

Influenza remains a leading cause of morbidity and mortality in human populations, causing 3 to 5 million severe infections and 290,000 to 650,000 respiratory deaths worldwide annually. Among influenza viruses, influenza A, including A(H1N1) and A(H3N2), and influenza B circulate and cause seasonal epidemics of disease [25]. Mortality caused by influenza has been aware of since the 1918 influenza pandemic [26]. Later, two more overwhelming pandemics, in 1957 and 1968, led to excess mortality, the difference between observed death and expected

death, in infants, the elderly and people with chronic diseases [27].

Thus, understanding the impact of influenza on mortality and providing correctly influenza forecasting is crucial to inform disease control decisions and decisions concerning the use of influenza vaccines [28]. Many researches have been conducted on forecasting influenza epidemics, retrospective analysis of mortality, and estimating excess mortality associated with influenza epidemics in Hong Kong and worldwide [29, 30, 31, 32]. Commonly used approaches can be concluded as compartmental models of infectious disease transmission (Susceptible-Infected-Recovered (SIR), Suceptible-Infectious-Recovered-Susceptible (SIRS), and Suceptible-Exposed-Infectious-Recovered-Susceptible (SEIRS) [32]) and statistical methods (Seasonal Autoregressive Integrated Moving Average (SARIMA), Poisson regression model, and log-linear regression model). Temporal features and the influenza surveillance data including influenza-like illness (ILI), laboratory-confirmed influenza positives, and severe cases have been used in almost all researches both as target and predictors. Other features selected by previous studies include demographic data for both population size and age group distribution, weather data, such as temperature and humidity, and mobility data for population movement patterns.

Deep learning networks have been adopted in recent years, such as self-attention-based network [33], feedforward neural networks [34], and long short-term memory (LSTM) model [35]. Compared to mechanistic approaches, deep learning methods can utilize more data besides surveillance data, such as protein sequencing data. Moreover, deep learning methods can capture the complex dynamics and interactions between the multiple influenza strains. Although those approaches have highlighted short-term performance, they usually have unsatis-

factory long-term accuracy and have limited interpretations of the influenza pandemic compared to compartmental methods and statistical methods.

1.2 Objectives and significance

This thesis aims to investigate the impact of COVID-19, influenza, and air pollution on mortality by 1) deploying statistical methods and machine learning models to estimate the key parameters of SARS-CoV-2 virus, understand the structure of excess mortality related to COVID-19, and have an insight into disease control measures and disease burden shifting; 2) using regression models and new metrics to evaluate whether the association between air pollution and mortality is consistent over time, especially during the COVID outbreak and updating air pollution impacts on specific-cause mortality; 3) exploring new variables in influenza activity forecasting and constructing neural network to balance model accuracy and interpretation.

1.3 Outlines

In Chapter 1, a background, objectives, and significance of this thesis were introduced and the outlines of the remaining thesis were discussed.

All-cause mortality during COVID-19 pandemic in Peru was analyzed in Chapter 2. We reported the peaks of all-cause mortality. Median age of daily death during the peaks was calculated and compared with that during normal period. Moreover, the primary cause of death for the direct and indirect excess mortality related to COVID-19 was examined. The structure of excess mortality related to

COVID-19 was revealed.

We first collected information on exposure history and symptom onset of Omicron BA.1 (i.e., B.1.1.529.1) cases and cases reported with Delta variants dominant in South Korea in Chapter 3. Then gamma distributions were deployed to estimate the incubation period. Both likelihood with interval censoring, and likelihood with convolution between Gamma distribution of incubation period of the assumed exponential distribution were applied. The maximum likelihood estimators of mean and standard deviation of those Gamma distributions were estimated with sensitivity analysis. The estimations of Omicron BA.1 were compared with the estimations during Delta variants dominance period. The findings were linked to disease control measures and gave an insight into disease control measures.

In Chapter 4, previous studies on the associations between air pollution metrics and mortality rates were reviewed. A new air pollution metric, "daily exceedance concentration hours" (DECH) introduced by [23] were reconstructed based on updated World Health Organization guidelines and the association between DECH and mortality risk across disease groups were examined in Hong Kong for pre-pandemic era as well as the pandemic period. The relative effects of DECH levels on mortality risk for different disease groups and model significance were reported. Whether the strength of associations between air pollution metrics and mortality rates are time-dependent was discussed. The change of association during the COVID pandemic was reported.

In Chapter 5, eXtreme Gradient Boosting (XGBoost) was used to estimate the disease burden attributable to influenza and environmental factors and compared with estimates from the general additive model (GAM) with a Gaussian link function. The weekly mortality rates of different age groups during pre-pandemic

(2014-2019) and pandemic period (2020-2021) were compared and discussed. Overall excess mortality, and excess mortality associated with influenza, air pollution was calculated. Possible reasons for the mortality changes were discussed.

In Chapter 6, previous influenza forecasting approaches were reviewed. The evolution of Human Influenza hemagglutinin (HA) and Neuraminidase (NA) sequences and their possible relationship between influenza outbreak and fatality were introduced. New deep learning framework was proposed to handle sequence data and time series data as well as providing certain interpretability.

Chapter 2

All-cause mortality during COVID-19 pandemic in Peru

Abstract: We reported the heterogeneity of the median age of all-cause mortality in Peru during different waves of COVID-19 pandemic. We believed that before the Omicron variants dominance, during the peaks of daily all-cause mortality, the median age of daily death was lower than the usual level. The median age of daily death bounced higher than normal during the peaks of daily all-cause mortality with the Omicron variants dominance. We also revealed the daily pattern of cause-specific mortality directly and indirectly related to COVID in Peru. We argued that most of the indirectly excess death in Peru were primarily caused by circulatory system diseases possibly caused by disruption in medical service, while the majority of directly excess death have the primary cause of death as COVID-19 and respiratory system diseases.

Keywords: COVID-19, Median age, Excess mortality.

COVID-19 has spread all over the world and by the end of 2021, the death toll had reached around 5.7 million with more than 384 million people being infected. Peru, as one of the most affected countries in the COVID-19 pandemic, has the highest number of confirmed cases, deaths per million, and total excess death [36].

Beaney et al. [37] suggested that distinguishing the direct COVID-19 and indirect COVID-19 death was crucial to understand the full impact of death caused by COVID-19. Previous studies conducted in Peru explored the direct and indirect impacts of COVID-19 on the wellbeing of population and qualities of healthcare system and delivery [38, 39]. However, there is limited research further classifying the direct and indirect COVID-19 death into specific cause of death.

Numerous studies mentioned that increasing age is strongly associated with COVID-19 mortality [37, 40]. The New York Times also reported that in the U.S. during the Omicron wave, older population had a higher mortality rate compared to previous waves [41]. On the other hand, few research discussed the heterogeneity of age distribution of daily death during the pandemic period in Peru.

We first used the death certificates dataset provided by Peru Ministry of Health to extract daily all-cause mortality and the median age of daily death from Jan 2019 to Apr 2022. Then we used the same dataset which listed every death registry with up to six descriptions of the cause of death. We identified a death is related to COVID-19 if at least one of the six descriptions of the cause of death included 'COV'. We sub-grouped cause-specific deaths based on the icd-10 code of the primary cause of death. The daily count of cause-specific death in Peru was visualized in several disease groups in comparison with the daily number of cause-specific death not related to COVID during the same period.

We noticed that during the COVID-19 pandemic before Jan 2022, there were

two peaks of all-cause mortality (Fig 2.1). During those two peaks, the median ages of daily death were lower than normal period, especially during the second peak roughly from the first half of 2021 with Lambda variant dominance according to GISAID [42]. After Jan 2022, the Omicron variants were dominant in Peru according to GISAID [42]. The peak of all-cause mortality was lower than the previous two peaks. However, the median age of daily death bounced higher to around 77 compared to 71 and 69 during the previous two peaks. One possible reason was that despite the lower fatality rate of Omicron variants, elderly population were more vulnerable than younger generation.

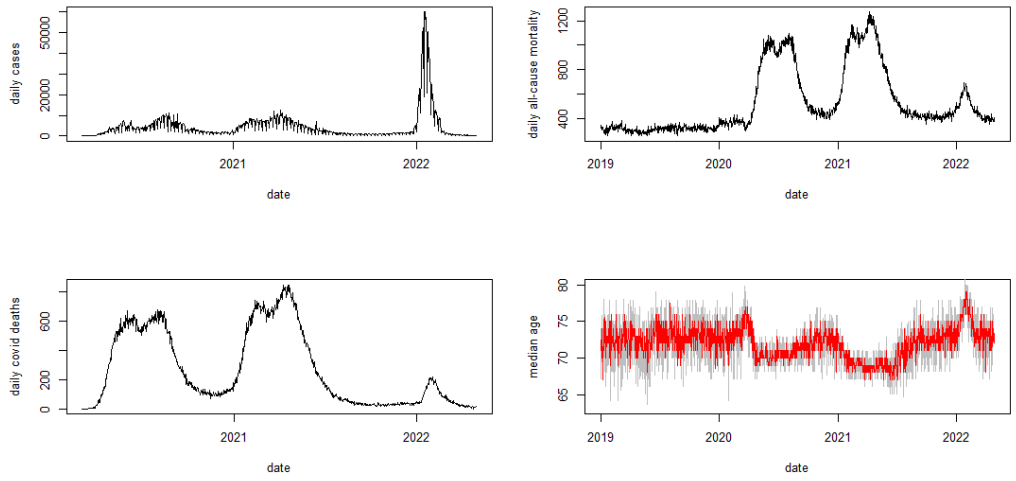

Figure 2.1: Daily confirmed COVID-19 cases in Peru (top left). Daily all-cause mortality (top right). Daily COVID-19 deaths (bottom left). Daily median age of all-cause mortality (in red), bootstrap 99% CI (shaded area) in Peru (bottom right).

We found a clear pattern of three peaks of total mortality counts during the three waves of COVID-19 (Fig 2.2). Compared to the pre-pandemic daily death counts, the excess mortality mostly fell into the following four disease groups: diseases of the circulatory system (ICD-10 codes starting with I), diseases of the respiratory

system (ICD-10 codes starting with J), general symptoms and sign (ICD-10 codes from R50 to R69), and COVID-19 (ICD-10 codes U71 and U72). Most excess deaths with primary cause of death as diseases of the respiratory system or COVID-19 (including virus not identified) were directly related to COVID-19 since the death registries had mentioned COVID-19 at least one time in the descriptions of the cause of death.

Most of the indirectly excess mortality had a primary cause of death as diseases of the circulatory system. During the first two waves of COVID-19, we can see a peak of roughly 100 daily excess deaths caused by circulatory system diseases. Among those indirectly excess deaths caused by circulatory system diseases, most of them had the cause of death listed as acute myocardial infarction (42.8%), cardiac arrest (20.7%), and heart failure (7.9%). Similar results have been found in Latvia [43]. It can be explained by the previous finding of the World Health Organization [44] stating that some countries have partially or completely disrupted services for cardiovascular emergencies during the pandemic. Furthermore, Peru's cardiovascular prevalence before the pandemic was also pronounced in the most urbanized regions, particularly on the Coast [45], overlapping with the areas hardest hit by the pandemic [46], which may have increased the unmet demand for cardiovascular services. Another explanation is that Coronavirus disease is associated with a high inflammatory burden which may cause cardiovascular disease [47]. Most of the excess mortality with general symptoms and sign listed as the primary cause of death could be concluded as a combination of both directly and indirectly related to COVID-19 based on the descriptions of all causes of death.

In summary, we reported the median age of daily death and the pattern of cause-specific daily mortality in Peru. We also revealed the directly and indirectly cause-

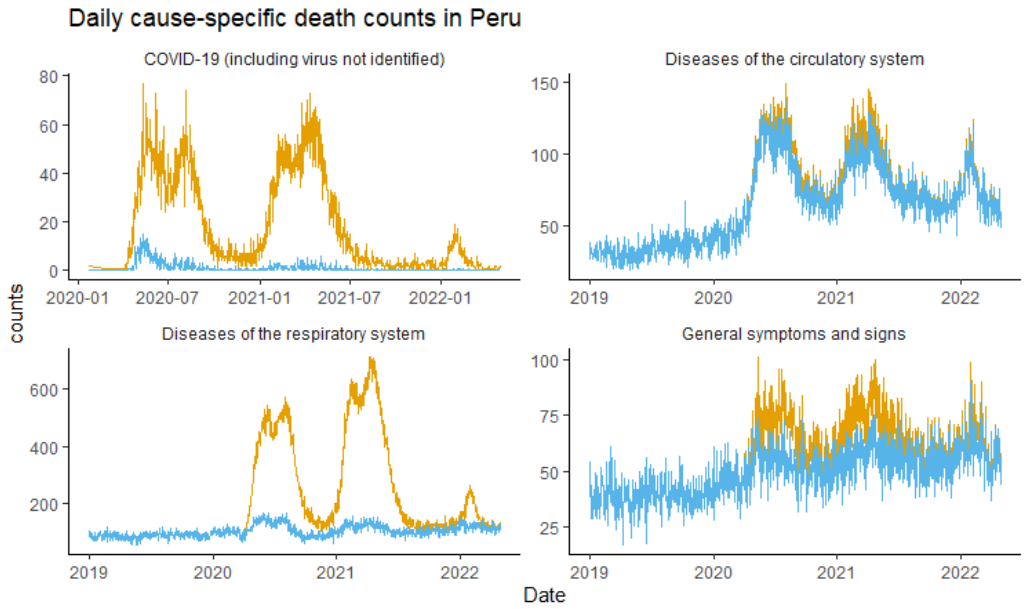

Figure 2.2: Daily death counts of four major groups of cause-specific mortality based on primary cause of death. The yellow lines include all death registries with cause of death information, while the blue lines exclude all death registries with 'COV' mentioned in at least one of the descriptions of the cause of death.

specific excess mortality during the three waves of COVID-19 by digging into all the descriptions of the cause of death of each deceased. We concluded that unlike the lower-than-normal median age of daily death during the first two waves of COVID, the median age bounced to a higher level when the Omicron variants were dominant in Peru. We figured that the majority of excess death directly related to COVID-19 had the primary cause of death as COVID-19 and respiratory system diseases, while most of the excess death indirectly caused by COVID-19 were primarily caused by circulatory system diseases, which may be the consequence of disruption in medical service during the pandemic.

Chapter 3

Estimating the incubation period of SARS-CoV-2 Omicron BA.1 variant in comparison with that during the Delta variant dominance in South Korea

Abstract: Based on exposure history and symptom onset of 22 Omicron BA.1 cases in South Korea from November to December 2021, we estimated mean incubation period of 3.5 days (95% CI: 2.5, 3.8), and then compared it to that of 6.5 days (95% CI: 5.3, 7.7) for 64 cases during Delta variants' dominance in June 2021. For Omicron BA.1 variants, we found that 95% of symptomatic cases developed clinical conditions within 6.0 days (95% CI: 4.3, 6.6) after exposure. Thus, a shorter quarantine period may be considered based on symptoms, or

similarly laboratory testing, when Omicron BA.1 variants are circulating.

Keywords: incubation period; Omicron BA.1 variant; Delta variant; quarantine.

3.1 Introduction

Since the end of 2019, COVID-19 has continuously posed threat to public health globally [48]. The novel genetic mutations of SARS-CoV-2 have continually challenged the control system for the COVID-19 pandemic, making it critical to monitor key epidemiological parameters for understanding the transmission and clinical characteristics of emerging variants [49, 50]. The incubation period is defined as the time interval between exposure and onset of illness for symptomatic infections [51], which is important to informing quarantine policies, to studying transmission dynamics of an infectious disease, and to assessing the effectiveness of entry screening [52, 53]. While estimates of incubation period can be found in literature for various historical SARS-CoV-2 strains [54, 55], the knowledge of incubation period for Omicron variants remains largely unassessed.

In this study, we collected information on exposure history and symptom onset of 22 Omicron BA.1 (i.e., B.1.1.529.1) cases in South Korea from November to December 2021, and estimated distribution of incubation period, which was then compared to that of 64 cases during Delta (i.e., B.1.617.2) variants' dominance in June 2021.

3.2 Method

3.2.1 Data collection

Based on the information of COVID-19 cases who tested positive for SARS-CoV-2 previously published [56, 57], we extracted exposure history and symptom onset date for patients with this information available. To use for incubation period estimation, we identified 22 cases laboratory-confirmed for Omicron BA.1 variants who were reported in South Korea from November 25 to December 31, 2021, and for comparison, we also included 64 cases reported in June 2021 when the Delta variants were dominant at a prevalence of 68.3% according to GISAID [42]. The exposure history was translated into exposure time window with upper and lower bounds of exposure date, which will be used for the calculation of the likelihood. Among these 86 (64+22) patients, all of them have illness onset date observed. Among the 22 identified Omicron BA.1 cases, 21 of them have both lower and upper bounds of exposure date, while 1 only has the upper bound of exposure date, and 12 cases during Delta dominance have both lower and upper bounds but 52 only have the upper bound.

3.2.2 Statistical analysis

Log-normal, gamma, and Weibull were among the most common distributions applied to estimate the incubation period [52]. The gamma distribution has a more concise mathematical expression compared to the other two distributions, hence less computational power is required to estimate the parameters. In this study, two different Gamma distributions ($f_{\text{incubation}}$) were adopted to govern the distributions

of incubation period for Omicron BA.1 cases and cases during Delta variants' dominance, respectively. For the samples with both lower and upper bounds of exposure date, i.e., with exposure window, we calculated the likelihood with interval censoring [52]. The exposure date was bounded within a time window, even though the exposure date could not be observed directly. We applied the interval-censoring on the likelihood function to account for the uncertainty of the observed exposure (or infection) time windows. Then, the difference between the observed illness onset date (denoted by S) and exposure date (denoted by I) of each individual case is the incubation period. As such, the likelihood function to estimate the incubation period was as follows,

$$L^{(1)}_{\text{incubation}} = \int_{I_U}^{I_L} f_{\text{incubation}}(S - I)dI$$

Here, I_L and I_U were the lower and upper bounds of exposure date, respectively.

For the remaining samples only with upper bound of exposure date observed, we assumed an exponential distribution indexed by this upper bound backwardly (denoted by function g), and calculated the likelihood with convolution between Gamma distribution of incubation period of the assumed exponential distribution [58, 59]. The likelihood function was as follows,

$$L^{(2)}_{incubation} = \int_{I_U}^{\infty} g(I - I_U) \bullet f_{incubation}(S - I)dI.$$

By using all samples, the overall likelihood is calculated by multiplying the two versions of likelihood functions. For the commonly noted selection bias issue of backward time interval during epidemic growth or decay phase [60], we consider

this is not applicable in our situations. For the cases infected by Omicron BA.1 variants, the epidemics curve from November to December 2021 presented an exponential growth with rate around 0.04 per capita per day, and thus we corrected the backward-observational sampling bias by using approach in previous study [61], and thus inferred the distribution of forward incubation period. For the cases collected in June 2021 when the Delta variants were dominant in South Korea, the epidemics curve appeared relatively flat in South Korea, which indicated an exponential growth with rate around 0, and thus there is unlikely to have selection bias due to backward observation.

We assumed the exponential infectiousness distribution has a mean of 3.7 days, which corresponded to the mean infectious period estimated in previous research [62]. We calculated the maximum likelihood estimators of mean and standard deviation of the Gamma distributions. We adopted non-linear optimization to maximize the overall log-likelihood function with a sufficiently small scale of $10^{(-6)}$ as relative tolerance level for convergence. To evaluate the statistical uncertainty, we used a parametric bootstrap with 1000 iterations of resampling to obtain 95% confidence intervals (CI) for each parameter. Limiting the dataset to those with exposure window observed, i.e., with both lower and upper bounds, we repeated the estimation with only 21 samples for Omicron BA.1, and 12 samples for Delta dominance period, respectively.

Sensitivity analysis was conducted by assuming shorter and longer versions of the exponential-distributed exposure window with 2.8 and 4.6 days to repeat the estimation (see Table in Appendices). Additionally, to relax the exponential assumption for the missing exposure window, we assumed the exposure windows of those samples only with upper bound of exposure date observed following an

empirical distribution from the samples with both lower and upper bounds of exposure date observed.

All analyses were conducted in R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria).

3.3 Results and discussion

For the 22 cases infected by Omicron BA.1 variants, the estimated mean incubation period was 3.5 days (95% CI: 2.5, 3.8), and SD was 1.4 days (95% CI: 1.0, 1.5), see Fig 3.1. We found that 50%, 95% or 99% of symptomatic cases may present clinical conditions within 3.3 days (95% CI: 2.4, 3.7), 6.0 days (95% CI: 4.3, 6.6) or 7.4 days (95% CI: 5.3, 8.2) after exposure, respectively. When limiting dataset to the 21 samples with exposure window observed, the mean incubation period decreased was estimated at 3.2 days (95% CI: 2.3, 3.8), see Table 3.3.

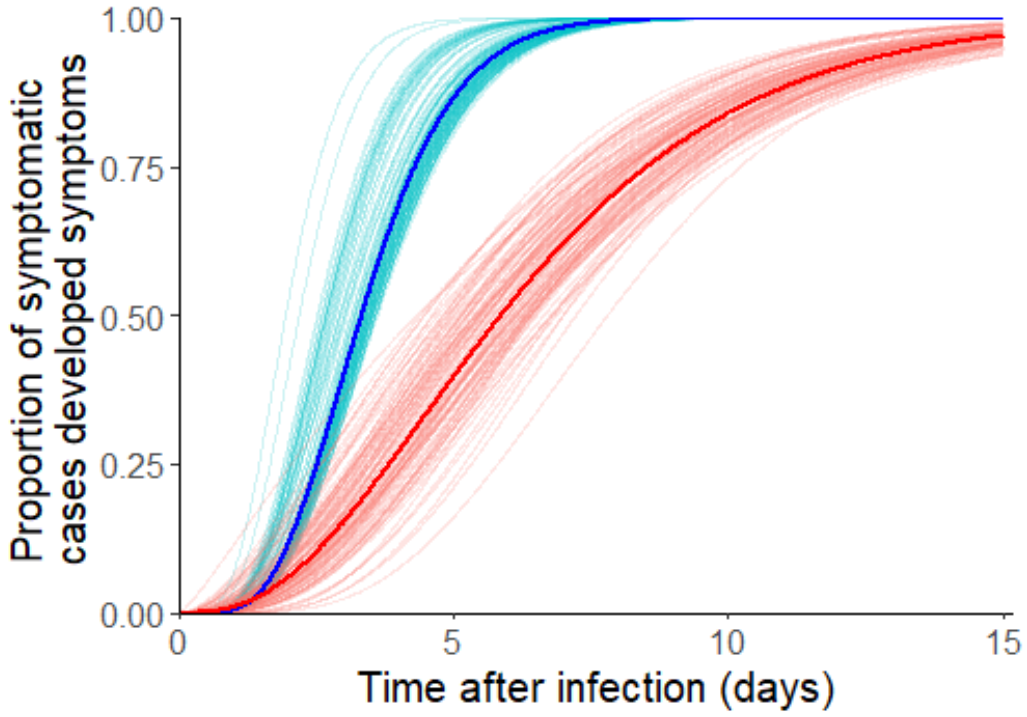

Figure 3.1: Estimated cumulative distributions of incubation period for Omicron BA.1 variants (in blue), and for cases during Delta dominance (in red). The statistical uncertainty was illustrated by 100 bootstrap estimates, which were curves in light colors, and the mean estimates were the bold curves in dark colors.

Table 3.1: Summary of incubation period estimates (unit: day) for cases infected by Omicron BA.1 variants and cases during Delta predominance period in South Korea.

Type of SARS-CoV-2 strain	sample[&]	sample size	estimate (95%CI)			
			mean	median	95-th percentile	SD
Omicron BA.1	all samples	$n = 22$	$3.5(2.5, 3.8)$	$3.3(2.4, 3.7)$	$6.0(4.3, 6.6)$	$1.4(1.0, 1.5)$
	with exposure window	$n = 21$	$3.2(2.3, 3.8)$	$3.1(2.2, 3.6)$	$5.5(4.0, 6.6)$	$1.3(0.9, 1.5)$
those during Delta dominance[$]	all samples	$n = 64$	$6.5(5.3, 7.7)$	$5.9(4.4, 7.1)$	$13.6(11.1, 15.9)$	$3.7(2.9, 4.6)$
	with exposure window	$n = 12$	$8.7(6.0, 11.6)$	$8.1(5.5, 11.0)$	$16.0(10.5, 21.0)$	$3.8(2.4, 5.6)$

Notes:
[$] These cases were collected in June 2021 when the Delta variants were dominant at a prevalence of 68.3% in South Korea according to GISAID [42].
[&] The samples "with exposure window" are those with both lower and upper bounds of exposure date observed, whereas "all samples" included the samples with exposure window and sample with only upper bound of exposure date observed.

By contrast, for the 64 cases identified during Delta dominance, the estimated mean incubation period was 6.5 days (95% CI: 5.3, 7.7), and SD was 3.7 days (95% CI: 2.9, 4.6). We found that 50%, or 95% of symptomatic cases may present clinical conditions within 5.9 days (95% CI: 4.4, 7.1), or 13.6 days (95% CI: 11.1, 15.9) after exposure, respectively.

For the sensitivity analysis, we found that the estimates with either shorter or longer version of exposure bound are consistent with main results in similar scales, which suggested the robustness of our findings, see Appendices. By using empirical distribution for those with missing exposure window, we found that the estimates were largely in line with the main results.

The mean and percentiles of incubation period of Omicron BA.1 variants were found considerably shorter than those of cases during Delta dominance period, as well as previous estimates based on other historical SARS-CoV-2 strains [54, 63]. Given the pre-symptomatic transmission feature of SARS-CoV-2 infection [64], a shorter incubation period indicated the Omicron BA.1 cases are likely to have a relatively higher rate at which they become new sources of infection to other susceptible individuals. Theoretical study also suggests that the generation time may be shortened with a short latent period [65], which is roughly equal to or less than the incubation period, and thus the Omicron BA.1 variants may lead to a lower period doubling time for epidemic curve regarding advantageous transmissibility in natural population and escape feature against herd immunity [66, 67].

Linking our findings to the disease control measures, some countries and regions have been using quarantine and entry screening as control measures against COVID-19. The initial quarantine periods were 14 days, and then extended to 21 days in some areas [68]. Although a longer quarantine period may lower the risk

of disease spread in community, people under quarantine or isolation were at risk of adverse mental health outcomes suggested by synthesized evidence [69], especially when the containment duration is longer than one week. Considering the latent period was typically shorter than incubation period [53], our estimates of the 99-th percentile at 7.2 days suggested a 7-day quarantine with PCR tests could be sufficient to detect around 99% of infections of Omicron BA.1 variants, and PCR tests have been confirmed effective to filter asymptomatic patients before they have onset of illness [70, 71].

There are some limitations in this study. First, for cases collected during Delta dominance period, we could not confirm these cases were infected by Delta variants due to the lack of genetic sequencing data. We could only conclude that the Delta variants were dominant at a prevalence of 68.3% in June 2021 in South Korea. Second, we adopted a Gamma distribution to govern the observed incubation period distribution, where symptoms were assumed to start immediately after infection. This may not be biologically reasonable, where a certain but minor lag may exist for patients to develop symptoms. Third, the exposure windows and illness onset time for patients can only be accurate to days. Therefore, a maximum of one-day error may exist in our determination of the intervals of exposure and symptom onset. Last, our estimate may be subjected to reporting and recall biases. It is suggested to further explore the heterogeneity of the incubation period among different SARS-CoV-2 Omicron variants, in order to adjust the disease control measures.

Chapter 4

Trends in the effects of ambient $PM_{2.5}$ concentration on mortality risk in Hong Kong, China

Background: Associations between levels of various types of airborne particulate matter such as ambient $PM_{2.5}$ and short-term mortality risk have been studied extensively. A metric called daily exceedance concentration hours (DECH) has been proved useful with respect to better modeling and understanding of acute mortality risk associated with pollution in southern Chinese cities. Notably however, it is unclear whether the strength of the association is time dependent. The current study investigated this using a comprehensive dataset acquired in Hong Kong spanning from 1999 to 2023. The methodology and modeling employed were similar to those used in prior studies.

Methods: Generalized additive models with quasi-Poisson distribution links were fitted to varying periods of an overall time series. These models were then exam-

ined to identify changes in implied effects on mortality risk over time.

Results: The replicated methodology of prior studies resulted in fairly consistent, but much reduced relative effects of DECH levels on mortality risk across the disease groups. The model remained significant with the inclusion of newer datasets. When applying the model to sliding time-windows of data, the effective risk of mortality remained relatively constant despite significantly changing levels of pollutants, especially with regard to mortality risk among cardiovascular diseases. Modelling other cause groups using DECH metrics yielded similar results to those acquired using other air pollution variables.

Conclusion: The results of the study support the use of DECH as a mortality risk factor, particularly with respect to cardiovascular diseases, and the size of the association is fairly consistent. During the COVID pandemic, the effect of DECH levels was reduced.

Keywords: air pollution, cardiovascular, Hong Kong, mortality, $PM_{2.5}$

4.1 Introduction

Numerous studies indicate that $PM_{2.5}$ is strongly associated with all-cause and specific-cause mortality [72, 73], but few reports mention whether the strength of associations between air pollution metrics and mortality rates are time-dependent. There are two possibilities. Either the size of an association is consistent, then one can use it with confidence to inform policymaking; or the size of an association is time-dependent, in which case identifying the mechanisms involved in variations would be informative. A recent study conducted by Lin et al. [23] introduced a new air pollution metric, “daily exceedance concentration hours” (DECH).

All conventional measures of air pollution concentration have trended down significantly in recent years (Figure 4.1). This includes the novel DECH metric. Intuition suggests that if DECH is a major indicator and cause of acute circulatory-cause mortality, as these levels decline over time the contributing risk of DECH should also decline. Most previous studies investigating air pollution and health hazards have focused on all-cause, circulatory-cause, or respiratory cause mortality, but in recent years more attention has been paid to mental, nervous system, and skin-related diseases [74, 75, 76]. Given that the quantitative association between these specific-cause mortalities and $PM_{2.5}$ is vague, a mathematical model using real-life data is necessary to fill the research gap.

Moreover, the outbreak of COVID-19 and the following disease control measures led to a significant decrease in all-cause and specific-cause mortality associated with air pollutants [77], and dramatic air pollution reduction including $PM_{2.5}$ in 2020 [78]. It's crucial to understand whether the association between specific-cause mortalities and $PM_{2.5}$ remain significant during the pandemic and post-pandemic period.

In the current study DECH and other variables were used to model all-cause and specific-cause mortality from 1999 to 2023 with time windows of different lengths, to investigate the sizes of associations between air pollution metrics and specific-cause mortality rates. The results of the study are organized into four sections; (A) replicating the methods of prior studies, (B) extending those methods to new data, (C) further exploring the conclusions of prior studies, and (D) applying models and the DECH metric to other diseases. The main target of this study is to investigate whether the strength of the association between $PM_{2.5}$ and mortality is time dependent for all-cause and cause-specific mortality.

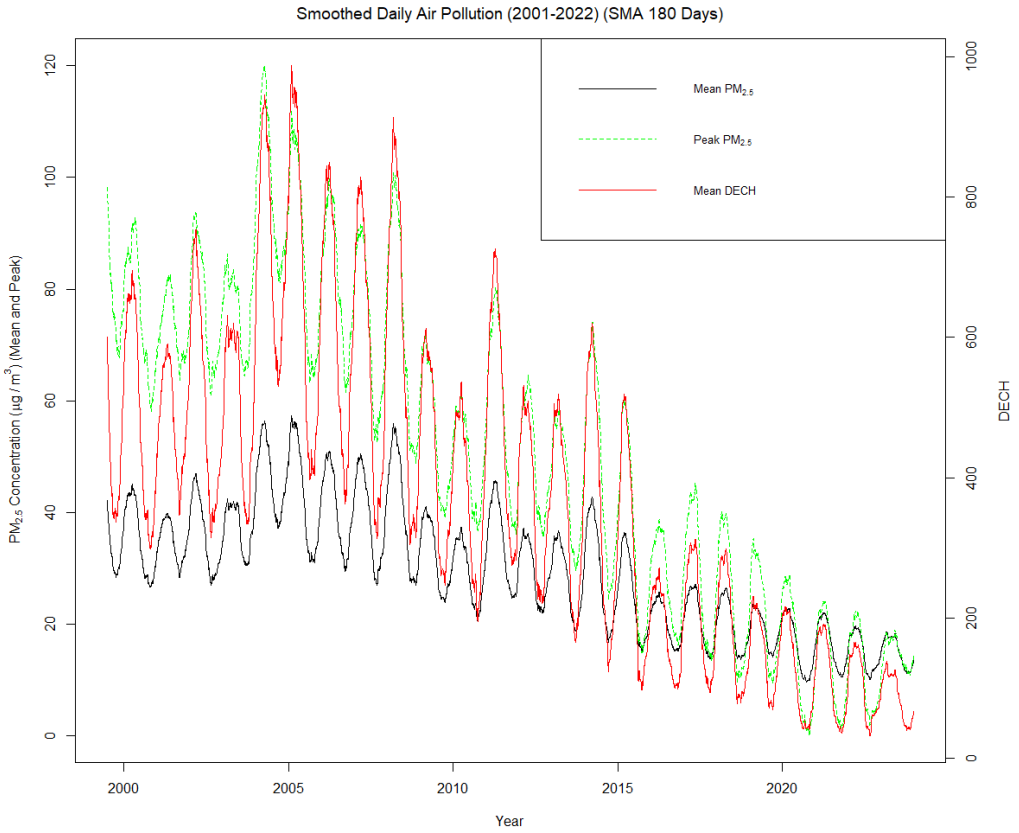

Figure 4.1: Simple moving average applied to daily mean and peak $PM_{2.5}$ concentrations and daily mean DECH indicating a trend of improved air quality in Hong Kong in recent years. Compared to the daily mean $PM_{2.5}$, the daily mean DECH better captures variations in air pollution within a day.

4.2 Methods

Data sources were used to gather daily information on mortality, air pollution, weather, hospital admissions for influenza and COVID-19, and public holidays in Hong Kong. All data were indexed daily to form a time series from 01 January 1999 to 30 December 2023. All data processing and analyses were performed using the statistical computation language R, and models were generated using

the ‘mgcv’ package.

4.2.1 Mortality

Mortality data were obtained from the death registry supplied by the Census and Statistics Department of Hong Kong. Data were filtered over three cause groups; all diseases, circulatory diseases, and respiratory diseases. Data from 01 January 1999 to 31 December 2000 were acquired, and cause of death was differentiated in accordance with the International Classification of Diseases (ICD) version 9. All deaths were filtered by numeric codes ranging from 001–799, deaths from circulatory diseases were filtered via codes 390–459, and deaths from respiratory diseases were filtered via codes 460–519. Data from 01 January 2001 to 31 December 2016 were differentiated in accordance with the ICD-10, therefore all deaths were filtered by numeric codes ranging from A00–R99, deaths from circulatory diseases were filtered via codes I00–I99, and deaths from respiratory diseases were filtered via codes J00–J99 [23]. Three more cause groups based on conditions commonly considered to be associated with air pollution were also incorporated into the current study; mental and behavioral conditions [74, 79], diseases of the nervous system and sense organs [80, 75], and diseases of the skin and subcutaneous tissue [76]. ICD-10 codes F00–F99 and ICD-9 codes 290–319 were used to filter deaths associated with mental conditions. ICD-10 codes G00–G99 and ICD-9 codes 320–389 were used to filter deaths associated with diseases of the nervous system. ICD-10 codes L00–L99 and ICD-9 codes 680–709 where used to filter deaths associated with skin diseases.

4.2.2 Air pollution

Hourly air pollution data including $PM_{2.5}$ levels were obtained from the Hong Kong Environmental Protection Department. Only 4 of 18 weather stations collected $PM_{2.5}$ levels before 2004, but more weather stations began to monitor $PM_{2.5}$ levels after that time. By the end of 2019 a total of 16 weather stations across Hong Kong were monitoring $PM_{2.5}$ levels. In the present study daily average pollution levels were calculated using all the data available for each given timepoint. In accordance with many prior studies [23, 81], daily mean and daily peak $PM_{2.5}$ concentrations were calculated. Daily meteorological data such as mean temperature (degrees Celsius) and relative humidity (percentage) were also collected. Daily data from all available stations where averaged to obtain daily means.

4.2.3 Influenza hospital admissions

Influenza hospital admissions data were obtained from the Hong Kong Department of Health's Centre for Health Protection. These data record the weekly influenza admissions totals. In accordance with [82] an "outbreak" week was defined as a week exceeding the 75th percentile of admissions for all weeks in that year. Notably the Centre for Health Protection has stated that "Since Feb 10, 2014, Public Health Laboratory Services Branch has adopted new genetic tests … this transition … may bring about increases in detection of and percentage positive for influenza viruses" [83].

4.2.4 COVID-19 surveillance data

Weekly surveillance data of COVID-19 were obtained from the Hong Kong Department of Health's Centre for Health Protection. Weekly fatal cases have been recorded since 2023. From 2020 to 2022, a separate dataset has been obtained from the Hong Kong Department of Health's Centre for Health Protection reporting daily covid death. Similar to the 'outbreak' week of influenza, an 'outbreak' week of COVID-19 was defined as a week exceeding the 75th percentile of fatal cases for all weeks in 2023.

4.2.5 DECH metric

As initially proposed by Lin et al. [84] the DECH metric is defined as 'daily concentration hours > 25 μg/m^3… [where] for example, an hour with a mean concentration of 28.5 μg/m^3 contributes 3.5 concentration-hours to the daily total; and hours with average concentration lower than 25 μg/m^3 contribute zero … to the daily total'. The boundary of 25 μg/m^3 was chosen by Lin et al. [84] based on guidelines published by the World Health Organization [18]. However, the guidelines published by the World Health Organization [19] updated recommended AQG levels to 15 μg/m^3. Thus, we redefined the DECH metric as 'daily concentration hours > 15 μg/m^3' in this research. DECH values were calculated for each day on a per-station basis, then the mean DECH of all available stations was used to define the DECH for that day over the region.

4.2.6 A general summary of the time series data

A general summary of the time series data mentioned in Section 4.2.1, 4.2.2, and 4.2.5 is provided in Table 4.1 below. From the summary statistics, it can be found that mental and behavioral conditions, diseases of the nervous system and sense organs, and diseases of the skin and subcutaneous tissue have limited mortality cases in Hong Kong. The spread of respiratory diseases related mortality is left skewed with relatively high variability. It may be caused by COVID related death.

Table 4.1: Basic information about the time series data

Variable	# Days	Mean ±-SD	Min	1st Q	Median	3rd Q	Max
Daily Mortality							
All diseases	9128	100.7 ± 28.65	44	84	96	112	481
Circ.	9128	21.33 ± 6.33	3	17	21	25	59
Resp.	9128	24.14 ± 17.77	3	16	22	29	347
Ment.	9128	2.18 ± 2.05	0	0	2	3	14
Nerv.	9128	0.89 ± 0.98	0	0	1	1	6
Skin	9128	0.50 ± 0.73	0	0	0	1	6
Air Pollution							
$PM_{2.5}$ DECH	9128	376.1 ± 406.9	0	64.5	230.8	570.9	3585.5
$PM_{2.5}$	9128	29.5 ± 19.2	2.6	15.4	24.6	39.9	172.0
Weather							
Temperature (°C)	9128	24 ± 5	5	20	25	28	32
Rel. Humidity (%)	9128	78 ± 10	27	73	79	85	99

4.2.7 Statistical model

A model was generated then applied to different segments of the time series data. In an effort to maximize consistency and reproducibility, a generalized additive model (GAM) with an expected quasi-Poisson distribution was generated in accordance with Lin et al. [84]. The aim of this model was to relate the discrete variable of daily circulatory mortality (count) to $PM_{2.5}$ concentrations. By finding

the coefficient on the DECH term for the model, a relative mortality risk effect percent relationship to changes in DECH $PM_{2.5}$ levels can be calculated. The model has been updated to consider the COVID-19 outbreak starting from 2020. The specific statistical model is as follows, where the time series Y is indexed by day, and hence $E[Y_t]$ gives the expected daily cardiovascular mortality at day t:

$$\begin{aligned}\log(E[Y_t]) =& \beta_1 \cdot \text{DECH}_{(-l)} + s(t, df = 6) + s(MT_0, df = 6) \\ &+ s(MT_{1-3}, df = 6) + s(MRH_0, df = 3) \\ &+ \beta_2 \cdot \text{INFL} + \beta_3 \cdot \text{COVID} + \beta_4 \cdot \text{DOW} + \beta_5 \cdot \text{PH} + \alpha\end{aligned}$$

DECH is the mean daily measure described in Section 4.2.5 for $PM_{2.5}$ concentration lag 3 days. DECH(-l) is lagged l day from t as described in [84], where acute mortality occurs between hours and days from initial exposure to elevated levels. MT is the mean temperature (degrees Celsius) at lag 0, and MT1-3 is a moving average of MT from days lag 1 through 3. This parameter was chosen for similar reasons as DECH being lagged 3 days. MRH is mean relative humidity (%) at lag 0. INFL is a dummy variable that takes the value of 1 when the given day at t is contained within a week designated as an 'outbreak' as described in Section 4.2.3 above. COVID is a dummy variable that takes the value of 1 when the givn day at t is contained within a week designated as an "outbreak" as described in Section 4.2.4 above. DOW refers to the day of the week, a dummy variable ranging from 0 to 6 from Monday through Sunday. PH is a dummy variable indicating a public holiday on the present day, where 0 indicates no holiday and 1 indicates a holiday (including Sunday, as defined by the Hong Kong government). The temporal index t was included to account for the clear trend and seasonality observed in Figure

4.1 above, and α is a random error term. The model incorporates smoother functions as penalized regression splines: s(). Degrees of freedom (df) were chosen in accordance with standards described in [84] and [85].

4.2.8 Model DECH lags

In the above model, DECH lag l was 2 days when applied to all mortalities, 3 days when applied to circulatory system mortalities, and 2 days when applied to respiratory system mortalities. These lag days were differentiated to match the significance figures identified and used by Lin et al. [84]. For the newly added cause groups, 1 day lag was applied to mental condition mortalities and nervous system mortalities, and 0 day lag was applied to skin mortalities [79].

4.2.9 Model objectives

The data sources and model were carefully constructed to replicate the methods described in [84] compiled with newly updated WHO air quality guidelines and the recent COVID pandemic. That study incorporated three models over the mortality groups; all cause, circulatory system, and respiratory system ranging from 1998–2011. The data used in the current study spanned from 1999–2023, facilitating testing and validation of the results over a more comprehensive scale. Three additional mortality groups were also incorporated into the current study; mental and behavioral, nervous system and sense organs, and skin. Notably the lack of 1998 data is due to fine suspended particulate (FSP) data not being available from the Environmental Protection Department for that year. It is unclear how other reports were able to include this data.

Part A of this study aimed to directly replicate results reported by Lin et al. [84] within the same time series with DECH threshold as 25 (for a fully reproduction), and Part B aimed to investigate validity beyond the fitted time series. In Part B the 13-year model in Part A was fitted on a sliding window basis starting in 1999, extending through 2007, and ending in years 2011 and 2023, generating 13 models to test the significance of the model on newer and out-of-sample data (data from 2012–2023). In Part C, to test shorter term changes in DECH, models were fitted to 5-year periods on a sliding window basis starting from 1999 and ending in 2023 inclusive, yielding fitted models across mortality groups for time series beginning with the year range 1999–2003, and extending to the year range 2019–2023. In Part D three additional models were incorporated, derived from the mortality groups mental and behavioral, nervous system and sense organs, and skin using 5-year periods on a sliding window basis starting from 1999 and ending in 2023 inclusive. This resulted in fitted models across mortality groups for time series with year ranges beginning at 1999–2003, and extending to 2019–2023.

4.3 Results

4.3.1 Replication of Prior Methods

For data ranging from 1999–2011, fitting the model described in Section 4.2 over all-cause, circulatory-cause, and respiratory-cause mortality groups generated the DECH coefficients (threshold '$> 25\ \mu g/m^3$' applied) shown in Table 4.2. The interquartile range (IQR) for hourly DECH measurements was 508.55 $\mu g/m^3$ throughout the period. This contrasts with the IQR of 565 $\mu g/m^3$ throughout

1998–2011 in [84]. The DECH coefficients were multiplied by this IQR to generate relative effect percentages for an hourly IQR increase in DECH concentration. Confidence intervals associated with the effect were determined by multiplying the standard error by 1.96. In Table 4.3 results reported by Lin et al. [84] are compared with results generated in the current study, including adjusted relative effect percentages using the Lin IQR value. Ratios of the current study's coefficients to Lin et al.'s [84] coefficients are also presented to neutralize any IQR issue by comparing ratios across groups.

Table 4.2: Fitted DECH coefficients for 1999–2011.

	Coefficient	Std. Error	Significance (Pr($>|t|$))
ALL DISEASES	1.763e-05	4.871e-06	0.000298
CIRCULATORY	2.825e-05	8.863e-06	0.00145
RESPIRATORY	2.155e-05	1.104e-05	0.0511

Table 4.3: Relative effect percentage comparison of fitted models.

Relative Effect %	All Diseases	Circulatory	Respiratory
Our Results (IQR 508.55)	0.90 (0.40, 1.39)	1.44 (0.53, 2.34)	1.10 (-0.03, 2.22)
Our Results (IQR 565)	1.00 (0.45, 1.55)	1.60 (0.59, 2.60)	1.22 (-0.03, 2.46)
Lin Results (IQR 565)	1.65 (1.05, 2.26)	2.01 (0.82, 3.21)	1.41 (0.34, 2.49)
Ratio (Our / Lin)	55%	72%	78%

4.3.2 Extending the model

To explore the validity of the model using other intervals and beyond the original sample data (1999–2011) a windowed approach was used to compute several models on a rolling basis. The values of the DECH coefficients (updated threshold '> 15 μg/m^3' applied) for a given window's model are shown in the following figures, with 0.05 significance level confidence intervals for each coefficient plotted above

and below in red. Fitting on 13-year intervals of data, the sliding window generated 9 models ending in years 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022 and 2023. Coefficient values for each mortality group are shown in Figure 4.2, and reference data for the plotted figures are presented in Appendices. During the pre-pandemic era, in the 13-year windowed models the DECH coefficients for all-cause and circulatory-cause cases reached significance at the 0.05 level for all windows, but respiratory models did not reach significance in the vast majority of cases. For the 13-year windowed models including pandemic era, the DECH coefficients started to lose significance for all-cause and respiratory-cause models, while the DECH coefficients of circulatory-cause models remained significance at the 0.05 level.

4.3.3 Extending the model to short-term intervals

Using a sliding window with 5-year intervals, models were fitted to identify short-term changes. DECH coefficients for each mortality group are shown in Figure 4.3, and reference data for the plotted figures are shown in Appendices. Multiplying each coefficient by the window's DECH IQR, the relative effect percentages across each mortality group are shown in Figure 4.4. Reference data for the plotted figures are shown in Appendices. After using a narrower time window, the vast majority of respiratory models remained lack of significance. For the all-cause and circulatory models, roughly half of them reached significance. Models that reached significance at the 0.05 level are indicated by "*" in Table 4.3.3. The relative all-cause mortality risk effect related to DECH fluctuated decreasing from roughly 1.5% to 0.5% across all models. The relative circulatory-cause mortality

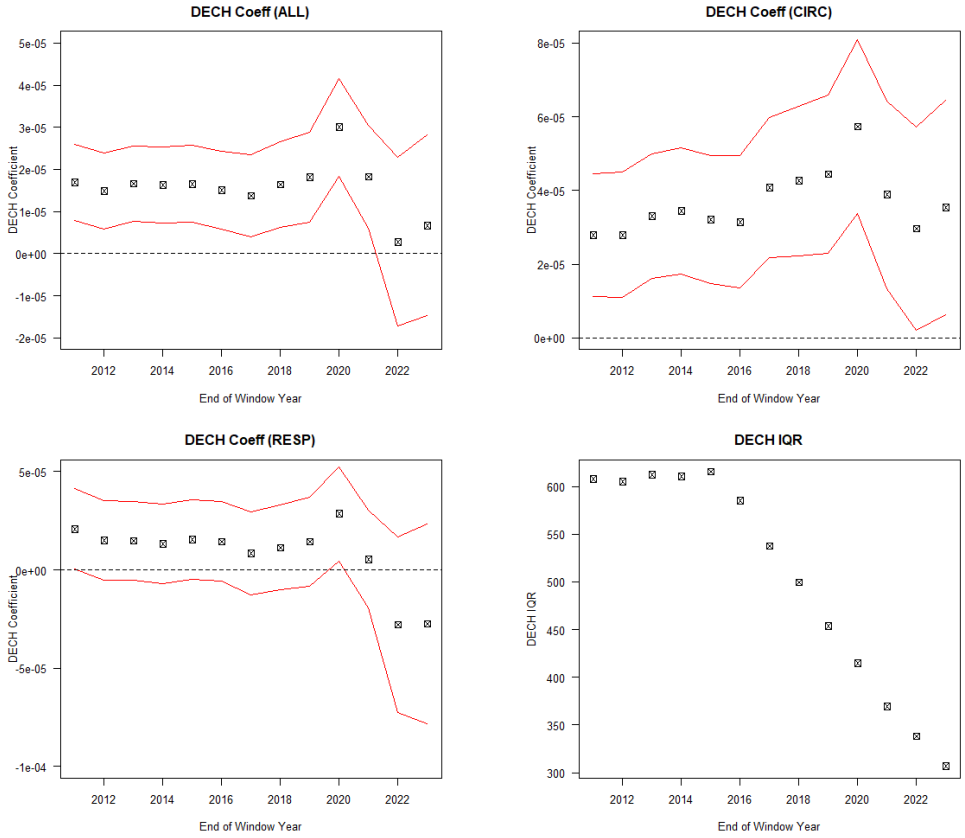

Figure 4.2: DECH coefficients across mortality groups for 13-year sliding window fitted models with 0.05 significance level confidence intervals. DECH IQRs associated with later sliding windows were dramatically lower, consistent with results shown in Figure 1. The DECH coefficients for all-cause and circulatory-cause groups reached the 0.05 level of significance for all windows in pre-pandemic era, while only the DECH coefficients for circulatory-cause group remained the 0.05 level of significance for windows including pandemic era. The respiratory models did not reach significance in the vast majority of cases. The DECH coefficients in circulatory-cause groups exhibited an increasing trend during the pre-pandemic period with a plunge during the pandemic period, while in all-cause groups they remained nearly unchanged during the pre-pandemic period with a similar plunge during the pandemic period.

risk effect related to DECH fluctuated around 1% across all models.

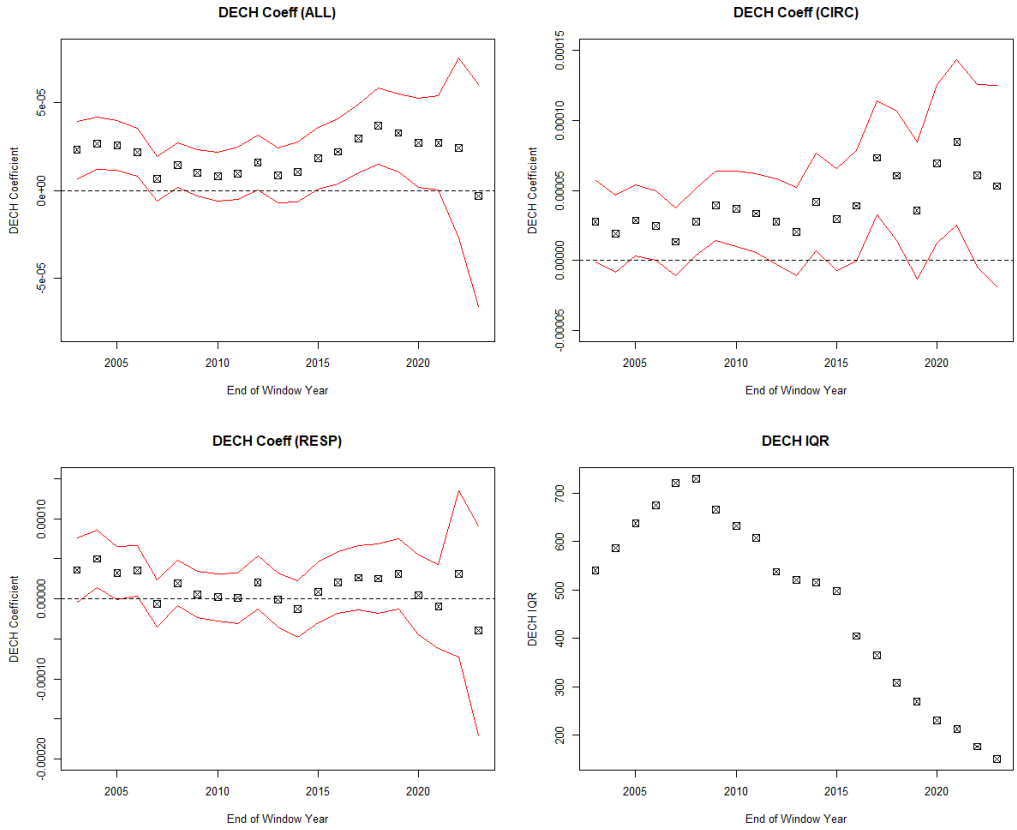

Figure 4.3: DECH coefficients across mortality groups for 5-year sliding window fitted models with 0.05 significance level confidence intervals. With a narrower time window the DECH IQR peak occurred during 2004–2007. The respiratory models did not reach significance in the vast majority of cases, whereas all-cause and circulatory-cause groups reached significance in some cases.

4.3.4 Applying the model to other cause groups

The same independent variables fitting the model described in section 2 were applied to data pertaining to mental, nervous, and skin diseases, which are commonly considered to be related to air pollution. A 5-year sliding window was applied to these models. The DECH coefficients for each window's model are shown in Fig

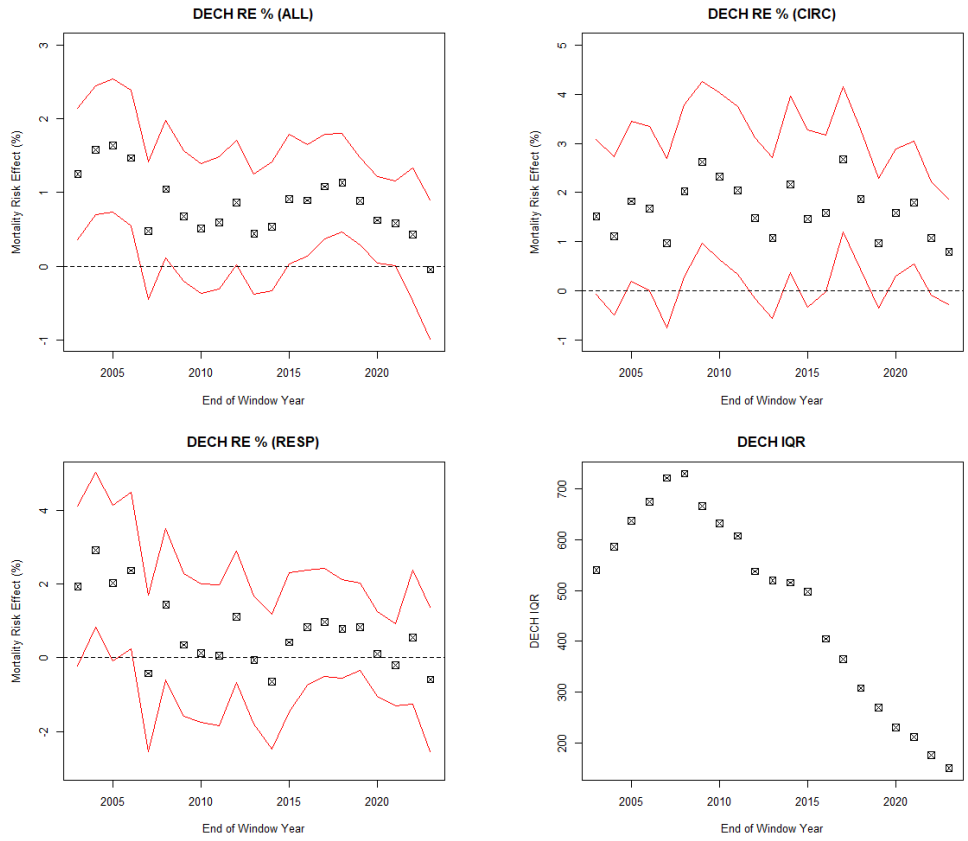

Figure 4.4: The relative effect percentages across mortality groups for 5-year sliding window fitted models with 0.05 significance level confidence intervals. The respiratory models did not reach significance in the vast majority of cases, whereas all-cause and circulatory-cause groups reached 0.05 significance in some cases.

4.5, with 0.05 significance level confidence intervals for each coefficient plotted above and below in red. Reference data for the plotted figures are shown in Appendices. From the generalized additive models applied to 5-year sliding windows for all mental-cause, nervous-cause, and skin-cause mortality groups, there was no clear clue of association between DECH variable and those cause groups.

Table 4.4: Indication of significance of each fitted model's DECH coefficient at the 0.05 level."*" indicates that the model term reached significance at the 0.05 level.

Window Year	All	Circulatory	Respiratory
2003	*		
2004	*	*	*
2005	*	*	*
2006	*	*	*
2007	*	*	
2008	*	*	
2009	*		
2010	*		
2011	*		
2012	*		
2013			
2014	*		
2015	*		
2016	*	*	
2017	*	*	
2018	*	*	*
2019	*		
2020	*	*	
2021	*	*	
2022			
2023			

4.4 Discussion

4.4.1 Application to New Data

Part A results in section 4.3.1 are generally consistent with [84]. All three mortality groups' models reached significance of the DECH term for their given lag, and the relative difference between mortality groups followed a similar pattern, i.e., the DECH coefficient for all-cause was much lower than that for circulatory-

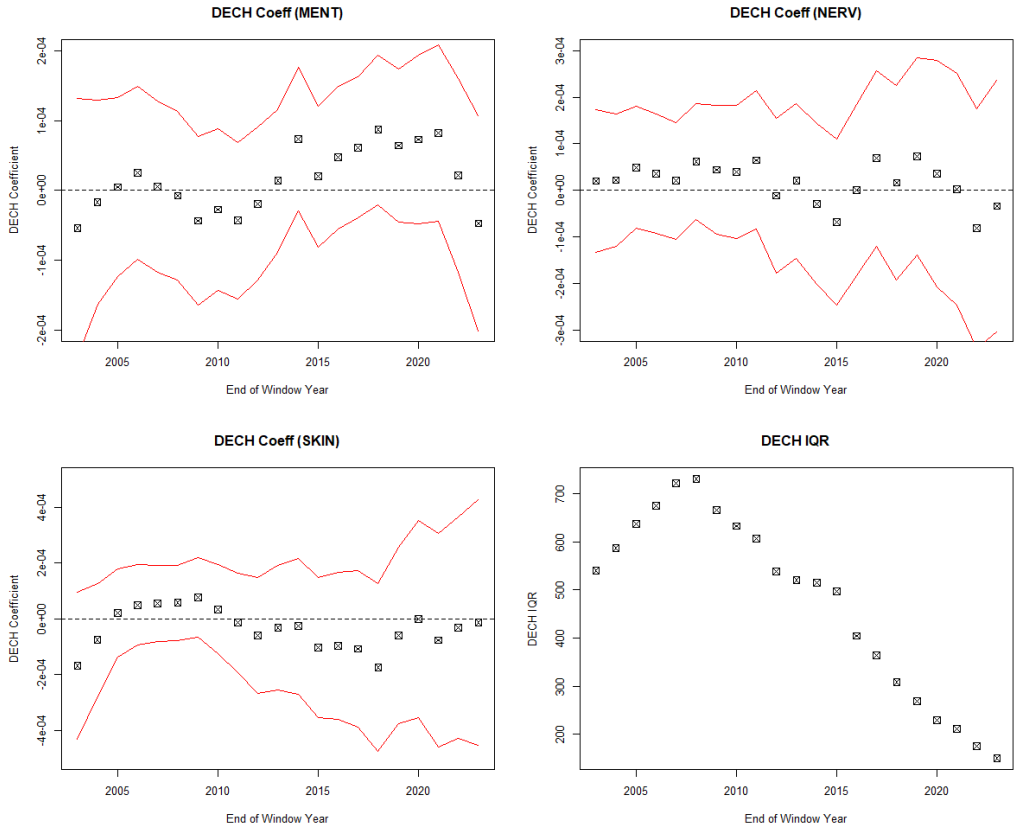

Figure 4.5: DECH coefficients across the new cause groups for 5-year sliding window-fitted models with 0.05 significance level confidence intervals. There were no clear associations between DECH level and nervous-cause mortality, mental-cause mortality, or skin cause mortality in any time windows.

cause, and respiratory-cause was somewhat lower than that for circulatory-cause but greater than that for all-cause.

While the circulatory and respiratory coefficient ratios were consistent with [84] (respective ratios of 72% and 78%), the all-cause coefficient was 55%, far less than the aforementioned circulatory and respiratory ratios. Further, all of the DECH coefficients indicated lower relative effect percentages, and it is unclear what the source of this large divergence between the two result sets could be be-

cause the vast majority of underlying data and modeling techniques used in the two studies were the same. The omission of 1998 weather data as described in section is a clear difference, however this is not believed to have had a strong effect due to the model's use of penalized splines to account for long-term trends.

4.4.2 Extending the model

When extending the model to years of data out-of-sample in Part A with a updated DECH threshold corresponding to the updated WHO guideline, significance of the DECH coefficients was reached for all-cause and circulatory-cause groups for all windows during pre-pandemic era. The coefficients in these groups were also fairly stable across windows before 2019, then a sudden increase in 2020 and followed by a plunge in 2021 and 2022. A possible explanation is that the association between air pollutants and mortality was covered by the surge of fatal cases of COVID in 2021 and 2022. Notably circulatory-cause coefficients exhibited an increasing trend during pre-pandemic era. This coefficient trend was neutralized by a rapidly lowered DECH IQR in recent years due to decreased air pollutants. Respiratory models did not reach significance in many cases, which is somewhat consistent with the weaker results reported throughout literature. Since we only have the data by the end of 2023, it's still unclear about the significance and effect of the DECH coefficients in the post pandemic period. It's crucial to keep attention to the association between mortality and air pollutants in Hong Kong during the post pandemic period considering a trend of improved air quality and a huge excess mortality in Hong Kong during the COVID pandemic which wiped out vulnerable population.

4.4.3 Extending the model to short-term intervals

The motivation to model based on a short-term interval such as 5 years was to investigate potential changes in relationships between FSP levels and mortalities in recent years compared to a decade prior where various measures of FSP clearly declined; principally from 2006 to 2023. In the present study most DECH coefficients derived from shorter-term respiratory models did not reach significance, whereas half of circulatory-cause and all-cause DECH coefficients were significant, especially those all-cause DECH coefficients derived from models fitted between 2010 to 2020. No models had significant DECH coefficients in the windows ending in 2007, 2012, 2013, or 2014. The DECH coefficients of all three models started to have a wider confidence interval after the COVID outbreak in 2020. Further investigation is needed to determine what led to this strong inconsistency.

Consistent with expectations, all-cause and circulatory-cause mortality were significantly related to elevated DECH levels, and there was an increasing trend in windows ending in years after 2011—which were above levels of the significant models ending in years 2008–2011. While circulatory-cause mortality exhibited a stable relative effect in later years despite increasing coefficients, this was due to simultaneously decreasing DECH IQRs, and is generally consistent with results presented by Lin et al. [84]. This indicates that DECH is a novel component of mortality risk, and the model presents a constant level of relative effect of DECH despite changing levels of pollution. The coefficients of DECH related to respiratory-cause mortality were fluctuated near zero, showing no evidence of relationship between FSP levels and respiratory-cause mortalities based on our generalized additive model.

4.4.4 Applying the model to other cause groups

In extensions of the model investigating three more disease groups, there were no significant DECH coefficients for any time windows with respect to diseases of the nervous system and sense organs, or skin and subcutaneous tissue. Those results are consistent with a previous study reported by Ho et al. [79] using Poisson regression models and hazy days as a predictor. For some recent time windows, there were some associations between DECH and deaths associated with mental and behavioral problems, which is consistent with a study reported by Ho et al. [86] that focused on associations between dementia mortality and environmental pollution. One possible reason for the association between DECH and mental-cause mortality in recent years is that society is paying more attention to mental and behavioral problems, and deaths related to mental health may have been misclassified in the past. In the mortality dataset there were more than 1000 deaths per year related to mental diseases after 2014, whereas there were less than 500 per year before 2009.

Compared to the circulatory-cause and respiratory-cause groups, which were included in the previous study, mental, nervous, and skin disease groups had lower daily death counts. The median of those three groups was to 1 or 0. Zero inflation can cause inaccuracy in quasi-Poisson models.

4.5 Conclusion

The current study builds on the incremental work of several researchers, principally that of [84]. Methods described by Lin et al. [84] were replicated, and

similar but much weaker effects of DECH levels on mortality risk were identified. There was also evidence to support the use of DECH as a mortality risk factor, specifically with regard to circulatory diseases. Despite the downward trend in air pollution in recent years, relative effects between mortality rate and DECH have remained stable. The COVID pandemic spread all over the world brought with large amounts of excess mortality and made the effect of DECH levels less significant for all disease groups. A further investigation of the relation between FSP levels and daily mortality in Hong Kong is recommended for the post pandemic era. The methods were applied to other disease specific mortality risks, and the results obtained were consistent with other studies.

Chapter 5

Change in disease burden associated with influenza and air pollutants during the COVID-19 pandemic in Hong Kong

Objectives: This study aimed to estimate the variation in disease burden associated with air pollutants and other respiratory viruses during the COVID-19 pandemic.

Methods: We adopted a machine learning approach to calculate the excess mortality attributable to air pollutants and influenza, during the pre-pandemic and pandemic period.

Results: In the first two years of the COVID-19 pandemic, there were 8,762 (95% confidence interval, 7,503 – 9,993), and 12,496 (11,718 – 13,332) excess all-cause deaths in Hong Kong. These figures correspond to 117.4 and 167.9 per 100,000

population, and 12.6% and 8.5% of total deaths in 2020 and 2021 respectively. Compared to the period before the pandemic, excess deaths from all causes, cardiovascular and respiratory diseases, pneumonia and influenza attributable to influenza A and B significantly decreased in all age groups. However, excess deaths associated with ozone increased in all age-disease categories, while the relative change of nitrogen dioxide (NO2) and particulate matters less than 10μm (PM_{10}) associated burden showed a varied pattern.

Conclusions: A notable shift in disease burden attributable to influenza and air pollutants was observed in the pandemic period, suggesting that both direct and indirect impacts shall be considered when assessing the global and regional burden of the COVID-19 pandemic.

Keywords: machine learning, mortality, COVID-19, influenza, air pollution

5.1 Introduction

As of 10 March 2024, the World Health Organization (WHO) reported that the COVID-19 pandemic has caused 775 million infections and 7 million deaths worldwide[2]. Prior to the availability of vaccines in early 2021, global health authorities adopted various strategies to curb the spread of the virus. Mainland China and the Hong Kong Special Administrative Region, which implemented a Zero-COVID policy to eradicate community outbreaks, enforced stringent public health and social measures. These included strict border control, mandatory quarantine for incoming travelers, and the promotion of social distancing through work-from-home policies and school closures [87, 88, 89]. Mask mandates were also enforced from July 2020 [90]. As a result, Hong Kong recorded a relatively

low number of COVID-19 cases and a low mortality rate in the first year of the pandemic, with only 148 deaths in 2020 [91].

The COVID-19 outbreak has significantly interfered the seasonal patterns of other respiratory viruses. Influenza activity, for instance, has remained low in many regions or countries since 2020, largely due to enhanced personal hygiene and social distancing measures [92]. The 2019-20 winter influenza season in Hong Kong, was remarkably shorter than previous seasons [93], which likely resulted in a significant reduction in disease burden attributable to the influenza virus. Prior to the pandemic, the WHO estimated that seasonal influenza was associated with 3 to 5 million cases globally, and 290,000 to 650,000 respiratory deaths annually [93]. In Hong Kong, our previous studies estimated that influenza resulted in approximately one thousand annual deaths and tens of thousands of hospitalizations [30, 94]. Most influenza associated deaths occurred in people with chronic diseases [95] and the elderly, who were also at higher risks of severe infections and deaths from COVID-19. Hence, the displacement of mortality or morbidity of susceptible populations by COVID-19 might have offset the disease burden caused by influenza. Similarly, improved air quality has been reported during the pandemic, likely due to social distancing and reduced economic activities [9]. While the direct effects of these containment measures on COVID-19 have been intensively evaluated [88, 96], few studies have investigated their indirect effects on the disease burden of other respiratory viruses and air pollution. A few studies estimated excess all-cause mortality in the pandemic across countries [97, 98]. However, most just calculated the difference between observed deaths and seasonal baseline based on simple time series models, few have considered other factors such as air pollution, temperature, and influenza in their models.

Previous ecological studies on the disease burden associated with influenza have often used time series modeling approaches such as quasi-Poisson regression or linear regression models [99, 100, 31]. These models share a common approach to calculating excess mortality. Initially, a model is fitted to the observed data and seasonal baselines are estimated from the model under the assumption that virus proxies are zero. The disease burden is then quantified by excess mortality, calculated as the difference between the observed data and the predicted baseline data. However, the selection of regression models and virus proxies remains a topic of debate. For example, the quasi-Poisson model with a log-link function has been used for weekly counts of all-cause and cause-specific mortality, but it has faced criticism for assuming a log-linear relationship between virus proxies and outcomes. The Gaussian linear model, which assumes a linear relationship between virus proxies and outcome variables, often yields negative effect estimates for virus proxies. These time-series models incorporate covariates such as seasonal trends, temperature, humidity, and proxies for influenza virus activities [99, 101].

In this study, we used a machine learning modeling strategy, eXtreme Gradient Boosting (XGBoost), to estimate the disease burden attributable to influenza and environmental factors. The XGBoost model offers the flexibility of assessing the marginal effect of each variable and minimizes overfitting through regularization penalties, bootstrapping of samples and cross-validation. This model has been applied to various research topics, including prediction models for disease incidence and prognosis, as well as the mortality burden of air pollution [102, 103, 104]. We also compared the XGBoost model estimates with those from the general additive model (GAM) with a Gaussian link function, which has been used in previous

studies on the disease burden of influenza [99, 105].

5.2 Method

5.2.1 Data

We obtained individual death records from 2014 to 2021 in Hong Kong from the Census and Statistics Department of the Hong Kong Special Administrative Region, China. These records were then aggregated to weekly death counts based on the International Classification of Diseases, Tenth Revision (ICD-10). As in our previous studies [30, 94], we considered the following categories: all-cause (ICD-10 codes A00-Z99), cardiovascular and respiratory diseases (CRD, ICD-10 codes I00-I99&J00-J99), and pneumonia and influenza (P&I, ICD-10 codes J10-J18). We further divided weekly death counts into six age groups: 0 -19, 20 - 39, 40 - 64, 65 - 84, and 85+ years. These age groups were selected to be consistent with our previous research on influenza disease burden [30]. Population data of these age groups were obtained from the Census and Statistics Department of Hong Kong. Weekly age-specific population size was calculated from annual mid-year age-specific population using a LOESS smooth function. Influenza surveillance data and weekly COVID-19 death counts were retrieved from the Centre for Health Protection. Meteorological data, including daily temperature and relative humidity, were downloaded from the Hong Kong Observatory (HKO) website. Air pollution data, including hourly concentrations of particulate matter less than 10 μm in diameter (PM_{10}), ozone (O_3) and nitrogen dioxide (NO_2), were retrieved from the 18 General and Roadside Air Quality Monitoring Stations of the Envi-

ronmental Protection Department (EPD).

5.2.2 Statistical Models

5.2.2.1 Model development

We first trained the XGBoost models on weekly death data by age groups. Weekly proportions of specimens testing positive for influenza type A (subtype H1N1 and H3N2) or type B were added to the models as proxy variables for influenza. The models also included weekly averages of mean, maximum, and minimum temperature, weekly average relative humidity, and weekly average concentration of PM_{10}, O_3, and NO_2. Other covariates in the models were week and year dummies to adjust for seasonal and annual trends. Weekly age-specific population was used as an offset in the model to adjust for population size.

Before fitting the models, weekly COVID-19 deaths were subtracted from weekly mortality data. We first trained the XGBoost model by repeatedly fitting models to the pre-pandemic data (2014-2019). We assumed different learning rates and maximum tree depth during the validation process with a 5-fold cross validation to prevent overfitting. The training procedure of the XGBoost models was conducted separately on all-cause mortality and cause-specific mortality by different age groups. The model parameters and number of boosted trees were decided based on the log-likelihood.

We also adopted a classical time series modeling approach, the GAM model with a Gaussian link function, which has been used in previous studies on the disease burden of influenza and air pollution. We fit the GAM models to all-cause mortality and cause-specific mortality, with adjustment for overfitting [106].

The best fit GAM model was selected after 5-fold cross validation. Cubic splines were applied to covariates like week number, average temperature, and relative humidity with 10 knots each. Similar to the XGBoost model, the GAM models also added weekly age-specific population size as an offset to adjust for age structure change over time.

5.2.2.2 Estimation of mortality burden

The mortality burden during the pre-pandemic and pandemic period was assessed using the following measurements: 1) Overall excess mortality in the pandemic: this was calculated as the difference between the observed death counts in 2020-2021 and the predicted counts from the XGBoost and GAM models using the observed data of covariates (meteorological data, air pollutants and influenza). 2) Excess mortality associated with influenza : We first estimated the influenza baseline mortality by setting the influenza proxy variable in 2020-2021 to zero (i.e., assuming no influenza virus activity) in the models. The difference between the observed death counts in 2020-2021 and the influenza baseline mortality was the excess mortality specifically attributable to influenza. 3) Excess mortality associated with air pollution : Using PM_{10} as an example, we first estimated the PM_{10} baseline mortality by setting the PM_{10} variable in 2020-2021 to its minimum value during the pre-pandemic period (9.7 $\mu g/m^3$), and subtracted this baseline data from the observed data to calculate excess mortality specifically attributable to PM_{10}. A similar calculation was repeated for O_3 and NO_2 (minimum values 23.8 and 13.8 $\mu g/m^3$, respectively). Considering the potential delay in mortality effects, we also calculated the lag effects up to 14 days prior for each influenza proxy and air pollutant. The 95% confidence interval (CI) for each estimate was calculated using

bootstrapping for 10,000 times.

All data analysis was conducted using the 'H2O' and 'xgboost' packages in R software (version 4.1.0) (R Foundation for Statistical Computing, Vienna, Austria). The R codes are publicly accessible at https://github.com/yanglin-polyu/covid-excess-mortality.

5.3 Results

5.3.1 Mortality burden during the pre- pandemic and pandemic period

Figure 5.1 shows a comparison of weekly mortality rates specific to different age groups during the pre-pandemic (2014-2019) and pandemic period (2020-2021). Mortality rates for all categories increased in the age groups of 20-39, 40-64 and 85+ during the pandemic compared to the pre-pandemic period. The numbers of all-cause deaths showed a sudden drop from February to April 2020, but gradually increased thereafter, peaking in early 2021 (Figure 5.2). Influenza nearly disappeared during these two years, with a small peak of influenza type B in winter 2020 only (Figure 9.1). Air pollutants generally remained at a relatively low level during the pandemic, with a sudden drop in early 2020 (except O_3), and gradually returned to pre-pandemic levels thereafter.

5.3.2 Overall excess mortality during the pandemic

Both the XGBoost and GAM models fitted well to the weekly data of all-cause mortality counts from 2014 to 2019 (Figure 5.2). The XGBoost models outper-

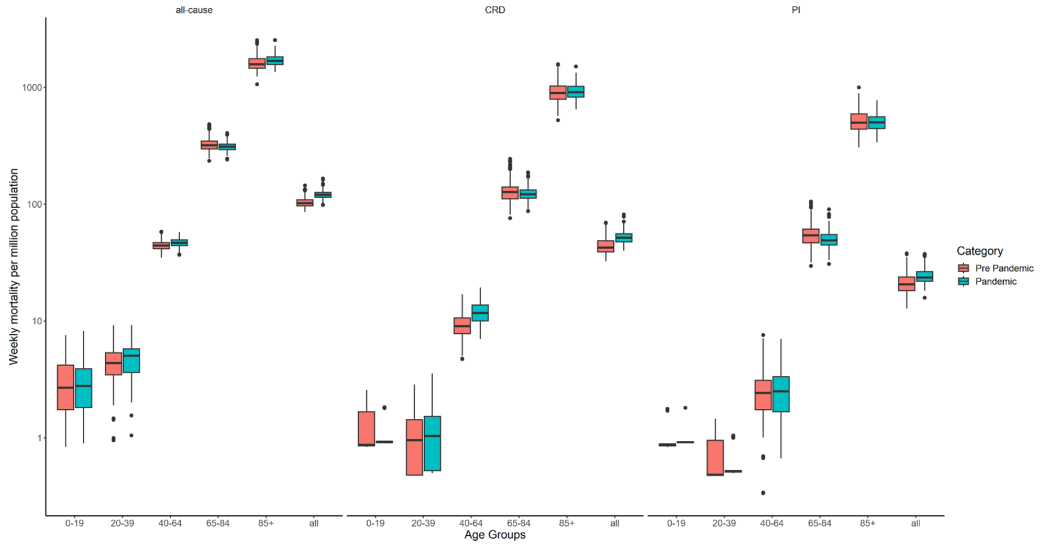

Figure 5.1: Weekly all-cause and cause-specific mortality by age group, during the pre- and pandemic period. CRD, cardiovascular and respiratory diseases; P&I, pneumonia and influenza.

formed the GAM models in terms of Model goodness-of-fit and prediction accuracy (Table 9.9). Given its capability of handling non-linear relationships and multicollinearity of predictors (Figure 9.2), we adopted the XGBoost models for main analysis and reported the estimates from these models hereafter. The greatest increase in excess mortality was observed in the 65-84 and 85+ age groups, with little to no significant increase in children and adolescents (Figure 5.2). A similar change in excess mortality was found in CRD, but not in P&I (Figure 9.3 and 9.4). There was a significant increase in all-cause excess mortality in Hong Kong during the first two years of the COVID-19 pandemic, with an estimate of 8,762 (95% CI, 7,503 9,993), and 12,496 (95% CI, 11,718 13,332) in 2020 and 2021, respectively (Table 5.1). These correspond to 117.4 and 167.9 per 100,000 population, 12.6% and 8.5% of total deaths in these years. The highest rate of excess mortality was found in those aged 65-84 years, with the excess rate estimates of

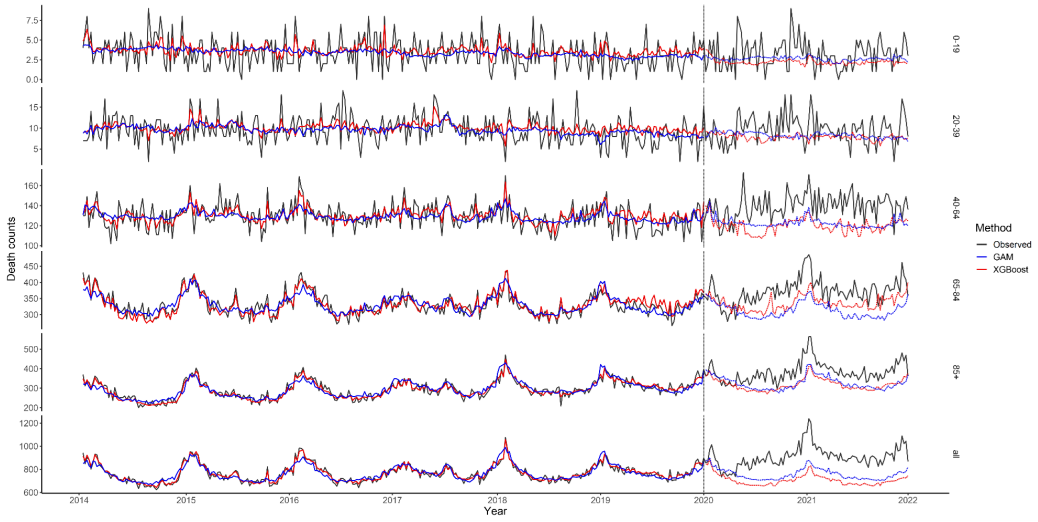

Figure 5.2: Observed and fitted weekly all-cause mortality data by age groups. The XGBoost (blue line) and GAM models (red line) were developed from the training data from 2014 to 2019 and used to predict the data in 2020 and 2021.

1,263.4 and 1,797.1 per 100,000 population, respectively (Table 5.1).

Table 5.1: Overall excess mortality numbers and rates (per 100,000 population) of all-causes, cardiovascular diseases (CVD), pneumonia and influenza (P&I) by age group in 2020 and 2021, estimated from the XGBoost models, respectively.

	All-cause		CVD		P&I	
	Excess number (95% CI)	Excess rate (95% CI)	Excess number (95% CI)	Excess rate (95% CI)	Excess number (95% CI)	Excess rate (95% CI)
			2020			
All-age	8762 (7503, 9993)	117.36 (100.17,134.71)	4221 (3571, 4874)	56.59 (47.82,65.36)	2199 (1874, 2530)	29.52 (25.22,34.08)
0-19	49 (19, 80)	4.48 (1.76,7.27)	-1 (-9, 9)	-0.07 (-0.82,0.79)	-1 (-6, 6)	-0.08 (-0.55,0.5)
20-39	108 (57, 159)	5.49 (2.9,8.07)	42 (19, 66)	2.12 (0.96,3.32)	1 (-7, 10)	0.06 (-0.36,0.52)
40-64	805 (590, 1018)	27.03 (19.63,34.14)	677 (557, 800)	22.63 (18.6,26.71)	123 (79, 169)	4.13 (2.64,5.67)
65-84	1399 (888, 1889)	118.65 (75.26,160.07)	989 (668, 1317)	84.31 (55.62,112.52)	431 (279, 587)	36.80 (23.45,50.62)
85+	2750 (2123, 3413)	1263.36 (978.04,1562.29)	960 (617, 1300)	440.54 (279.09,598.14)	672 (465, 887)	311.33 (214.32,405.93)
			2021			
All-age	12496 (11718, 13332)	167.86 (157.43,179.48)	6008 (5539, 6534)	80.60 (74.29,87.56)	2911 (2675, 3150)	39.07 (35.93,42.27)
0-19	29 (8, 50)	2.66 (0.74,4.63)	-5 (-11, 1)	-0.47 (-1.04,0.12)	0 (-4, 5)	0.04 (-0.33,0.49)
20-39	111 (66, 156)	5.75 (3.5,8.01)	37 (21, 54)	1.93 (1.07,2.84)	2 (-6, 11)	0.12 (-0.32,0.59)
40-64	1150 (976, 1325)	38.52 (32.7,44.17)	811 (699, 919)	27.02 (23.38,30.65)	103 (59, 147)	3.43 (1.96,4.91)
65-84	1770 (1325, 2221)	146.33 (109.4,182.39)	1312 (1086, 1552)	107.96 (88.82,127)	412 (271, 546)	33.87 (22.47,45.15)
85+	4062 (3650, 4480)	1797.08 (1618.03,1982.17)	1518 (1226, 1815)	672.23 (544.9,803.74)	973 (792, 1165)	431.59 (346.89,512.62)

5.3.3 Excess mortality associated with influenza

During 2014-2019, influenza A (H1N1 and H3N2) resulted in more deaths than influenza B in children and adolescents, but fewer in older adults (Figure 5.3). A similar pattern was observed in CRD and P&I, except for more P&I deaths associated with influenza A in the 85+ group (Figure 9.5 and 9.6). Annual excess mortality rates attributable to influenza A and B dramatically decreased in all age groups during the pandemic period, compared to the pre-pandemic period (Figure 5.4). The largest decrease (87% reduction from pre-pandemic estimates) was observed in excess mortality associated with PM_{10} in the 0-19 group, and the highest decrease (273.47 times higher) was observed in influenza B associated deaths in the 85+ group. Similarly, annual rates of CRD and P&I excess mortality attributable to influenza A and B dramatically decreased in all age groups during the pandemic (Figure 9.7 and 9.8).

5.3.4 Excess mortality associated with air pollution

During the pre-pandemic period, annual all-cause excess mortality associated with NO_2 and PM_{10} showed a decreasing trend across years in all age groups, while O_3 estimates slightly increased (Figure 5.3). Annual all-cause excess mortality rates attributable to air pollutants decreased in all age groups during the pandemic period, with a few exceptions including O_3 estimates in the all-age group (Figure 5.4). The estimates of CRD and P&I excess mortality rates attributable to all air pollutants had a similar pattern (Figure 9.7 and 9.8).

We also conducted a sensitivity analysis by incorporating data on air pollutants and influenza proxies from up to 14 days before (Figure 9.9). The model estimates

are generally similar to the main results. Compared to the XGBoost models, the estimates from the GAM models were more conservative than those from showing negative values in several age-disease categories. However, the overall trends of relative change remained consistent between the two models (Table 9.10).

5.4 Discussion

In this study, we observed a significant rise in all-cause, CRD and P&I mortality during the pandemic in Hong Kong. The XGBoost model estimated excess mortality at 21,258, 10,229 and 5,110, while the GAM model estimated it at 15,431, 8,426 and 2,666, respectively. These numbers are significantly higher than the 148 COVID-19 deaths reported during the same period. As expected, older age groups (65-84 and 85+ years) had a greater increase than younger age groups. Interestingly, the annual all-cause excess mortality rates attributable to both influenza and three common air pollutants decreased in all age groups during the pandemic compared to the pre-pandemic period, with the exception of a slight increase in all-cause excess mortality rates associated with O_3. Although we included time, temperature, relatively humidity, respiratory viruses, and air pollution as covariates in the models, there were likely other unadjusted factors contributing to the increased mortality burden. We hypothesize that changes in health seeking behavior could be one potential reason, with people possibly delaying diagnosis and treatment due to fears of contracting COVID-19 in healthcare facilities. This is supported by reports of a significant decrease in overall hospitalizations in Hong Kong in 2020, with a corresponding decrease in in-hospital mortality and an increase in out-of-hospital mortality [107].

Most previous studies on excess mortality only considered time and seasonal trends in predicting baseline deaths, and some even produced negative estimates[37]. This highlights the difficulty in assessing the true impact of the pandemic. Our study is one of the first to conduct a comprehensive investigation into the relative change in disease burden associated with different factors, including influenza and air pollutants. It is not surprising to observe a dramatic drop in all-cause excess mortality rates attributable to both influenza A and influenza B during the pandemic. This is likely due to enhanced personal hygiene and social distancing measures, which have been reported to have nearly eliminated other respiratory pathogens during the COVID-19 pandemic in many countries and regions [108, 109]. Additionally, the uptake rates of seasonal influenza vaccination increased in the Hong Kong population, particularly among children aged six months to six years, with rates increasing from 19.2% in 2016/17 season to 47.4% in 2019/20 season [110].

Previous studies have investigated changes in air pollutants during lockdowns. In China, it was found that NO_2 and CO concentrations decreased during the lockdown period, while O_3 increased. Based on these three air pollutants, the lockdown policy prevented certain all-cause deaths [111]. A global study involving air quality stations from 34 countries estimated that a net total of 49,900 excess deaths were avoided during lockdowns due to reduced emission of NO_2, O_3 and PM2.5. In China, the $PM_{2.5}$-related avoided excess mortality was 19,600 [112]. However, studies in England and Australia found no association between air pollution and excess mortality when comparing the pre-pandemic and pandemic periods [113], and the change in air quality during the COVID-19 lockdown had a negligible impact on calculated health outcomes [114]. Whether the reduction in air

pollution during the lockdown has an effect on total mortality remains a question, but studies have found that air quality is the most important factor in the context of enabling an increase in COVID-19 case load [115]. A review by Becchetti et al [116] found a strong association between long-term air pollution exposure and COVID-19 deaths. In our study, we found that the annual all-cause, CRD, and P&I excess deaths attributed to NO_2 and PM_{10} decreased during the pandemic in Hong Kong based on our XGBoost model, while the excess deaths attributed to O_3 increased slightly, similar to the study in China [111]. We also observed an increase in the excess deaths attributed to O_3 in the pre-pandemic period. In addition to the excess deaths attributed to air pollutants, we found a significant decrease in the annual all-cause, CRD, and P&I excess deaths attributed to both influenza type A and type B. Although the predicted all-cause weekly mortality counts for each age group during the pandemic period were lower than those in the pre-pandemic period due to the decrease in excess deaths attributed to major air pollutants and influenza proxies (with the exception of excess deaths attributed to O_3), there was a significant amount of all-cause and CRD excess mortality counts for most age groups based on both prediction methods after deducting the weekly COVID-19 deaths.

Few studies have simultaneously assessed the disease burden of influenza and air pollutants, primarily due to the multicollinearity between these variables (Figure 9.2). This challenge is effectively managed by our XGBoost models, which utilize decision trees and regularization techniques to deal with non-linear relationships. While accurately determining the disease burden from influenza and air pollution is challenging, our calculated excess mortality for the period before the pandemic align well with findings from earlier research. For instance, a study

conducted in China in 2017 estimated around 4,200 deaths (with a range of 3,300-5,100) in Hong Kong were due to air pollution [117]. Another study estimated around 3,000 deaths in Hong Kong in 2013 were attributable to air pollutants [118]. Regarding the burden from influenza, estimates from previous studies using GAM models suggested between 500 and 1,000 excess deaths from influenza A and B in Hong Kong between 1998 and 2009 [30, 99]. These numbers are within the range of our XGBoost model estimates, which suggest between 4,500 and 5,000 excess deaths from 2014 to 2019 due to these factors (Figure 5.3). Taken together, these findings support the reliability of the XGBoost models in providing estimates for the disease burden from air pollutants and influenza.

There are some limitations in this study. First, we only retrieved meteorological data, air pollution data, and influenza data. We did not include any demographic information. Second, we chose the GAM model with a Gaussian link function, which may have negative effect estimates of influenza, making it difficult to compare the influenza and air pollution associated mortality burden calculated with different models. Finally, further research should be conducted to explore the possible reasons for excess mortality in Hong Kong during the pandemic period.

5.5 Conclusion

Using advanced machine learning approaches, we estimated a significant decrease in disease burden associated with influenza and air pollutants in a region with minimal COVID-19 cases in the first two years of the pandemic, while the overall mortality burden during the pandemic period increased compared to the pre-pandemic period. Our findings suggest that comprehensive assessments of the global and re-

gional burden associated with the COVID-19 pandemic should consider its direct and indirect impacts.

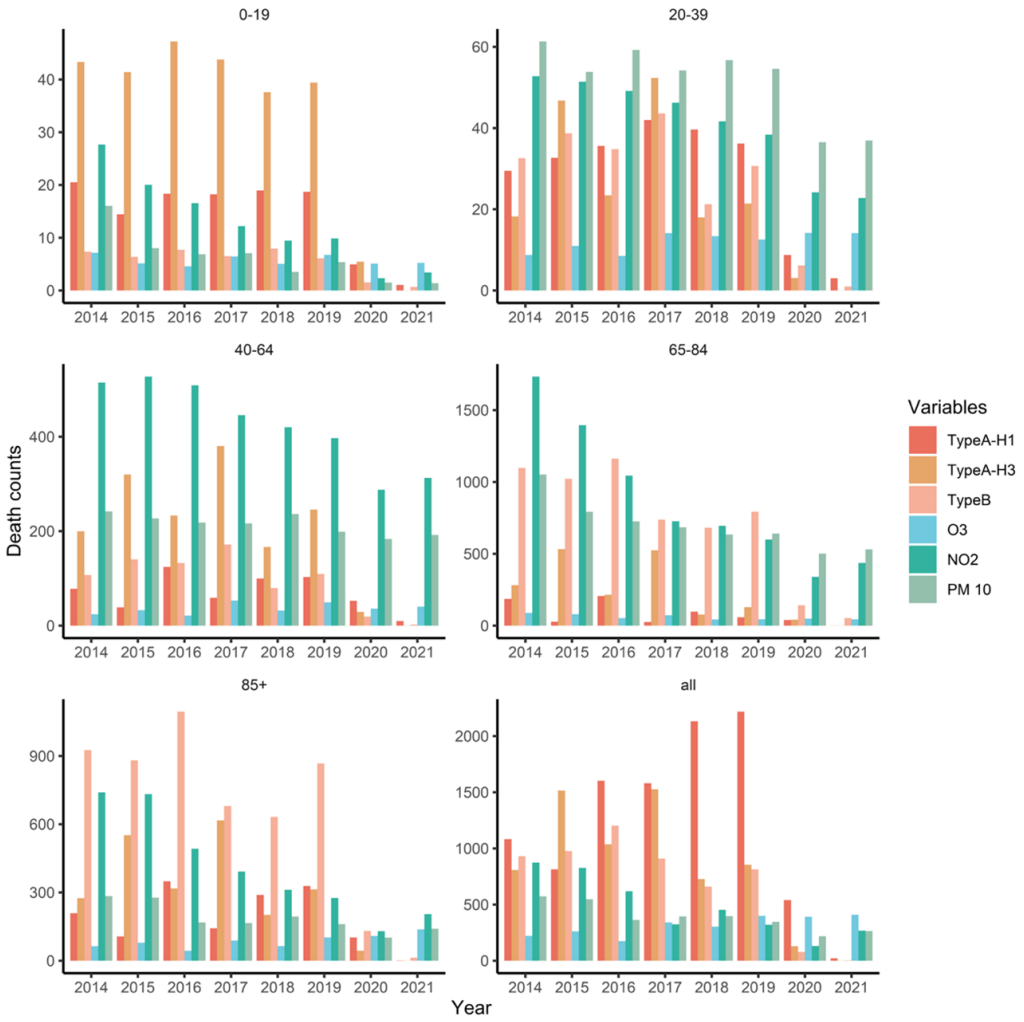

Figure 5.3: Annual all-cause mortality associated with influenza A subtype H1N1 (red bar), H3N2 (orange bar), influenza B (pink bar), O_3 (blue bar), NO_2 (green bar), and PM_{10} (light green bar). The estimates were derived from the XGBoost model.

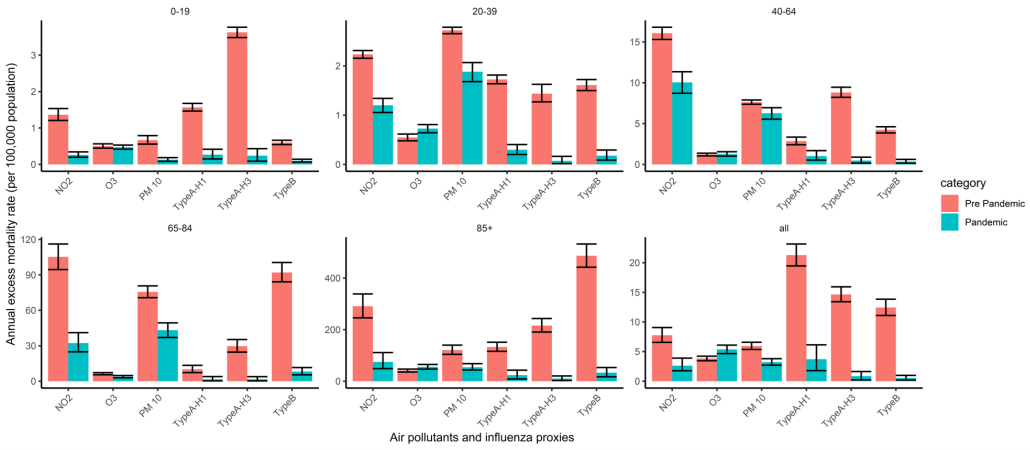

Figure 5.4: Comparison of annual all-cause excess mortality rate (per 100,000 population) associated with air pollutants, influenza A and B between the pre- and pandemic period, for different age groups, with 95% confidence interval (error bar) obtained from bootstrapping for 10,000 times.

Chapter 6

Future Study: Forecasting long-term influenza activity in Hong Kong

6.1 Background

An accurate forecast of the influenza activity can help to make public health interventions [119]. Beginning in 2013, the U.S. Centers for Disease Control and Prevention (CDC) launched FluSight, a forecasting initiative, to decrease the uncertainty of annual impact of flu at the start of each flu season. The primary targets of the FluSight challenge included weekly incidence rates, season onset, peak timing, and peak intensity [120]. Various research groups and modeling teams used either statistical time-series models or mechanistic transmission models and historical data from surveillance systems to predict key features of seasonal influenza [32].

However, unlike the highly seasonal epidemics in temperate regions (like US), Hong Kong, a subtropics city, can have influenza epidemics occurred at any time of the year (Figure 6.1) and, often, multiple types/subtypes co-circulate [121]. Many studies have been conducted in Hong Kong on influenza forecasting using either traditional statistical method [34, 122] or compartmental model [29] similar to the study in US. Instead of forecasting key epidemiological indicators, most of them focused on the time series prediction of weekly influenza-like illness consultation rate. Yet, those studies focused on short-term or even one-week-ahead forecasts. Considering the reporting lag of influenza surveillance data, short-term forecasts may provide limited insights for real-world epidemic management. A long-term forecast of influenza activity can help decision-makers to design comprehensive intervention strategies such as school closures, community-wide social distancing measures, or public awareness campaigns.

After the outbreak of COVID-19, the influenza activity remained in a low level (Figure 6.1). [109, 123] discussed the reduction in influenza virus infection and possible reasons included changes in human behavior and the widespread implementation of nonpharmaceutical interventions (NPIs) such as mask mandates, social distancing, travel restrictions, and enhanced hygiene practices. Besides the reduced influenza transmission caused by the NPIs deployed to manage COVID-19, the respiratory viral testing power has been enhanced dramatically [12] and possibly created bias when used influenza surveillance data, such as weekly positive detections of seasonal influenza virus and weekly positive percentage for influenza among respiratory specimens, before and after the COVID pandemic. Weekly number of severe influenza cases and the weekly consultation rates of influenza-like illness reported by General Out-patient Clinics (GOPC) can better

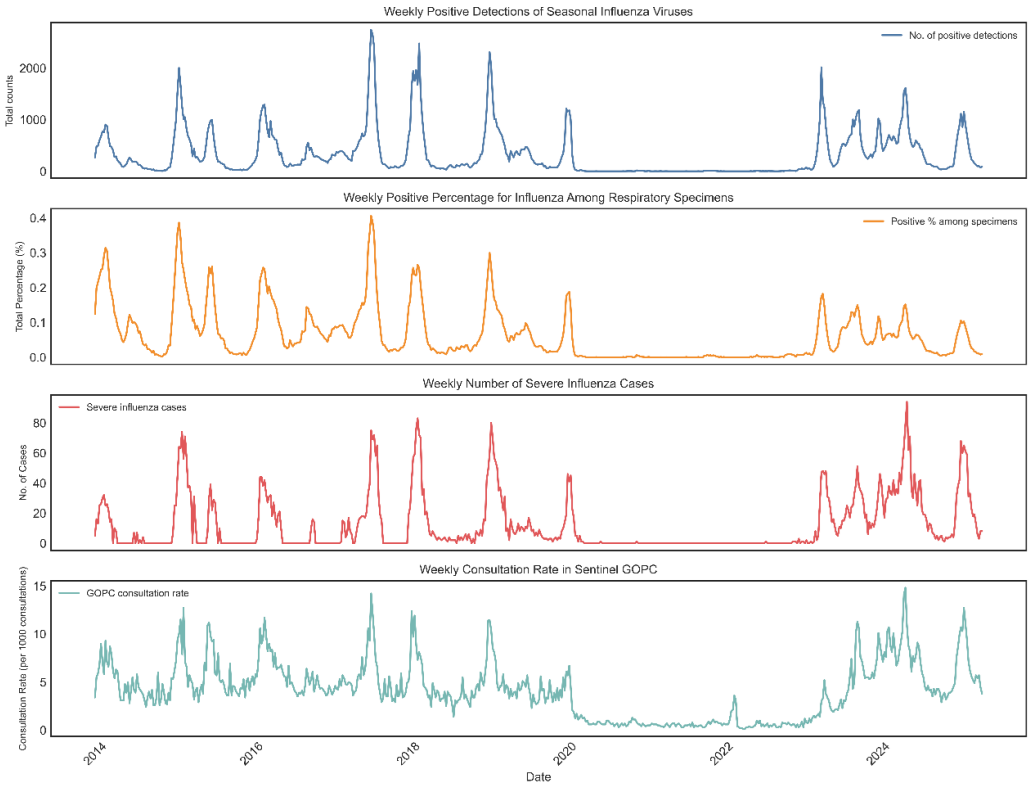

Figure 6.1: Comparison of different weekly influenza indicators over 10 years in Hong Kong. Weekly positive detections, positive percentage for influenza, number of severe cases, and influenza-like-illness consultation rate in GOPC appeared similar pattern before the COVID-19 pandemic. In the post-pandemic era, the maximum values of weekly positive percentage for influenza among respiratory specimens remained considerably lower compared to the pre-pandemic period.

capture the trend of influenza activity and maintain consistency over time.

A key aspect of seasonal influenza outbreaks is that the influenza virus could evolve over time and evade the immune system [124]. The influenza virus's hemagglutinin (HA) and neuraminidase (NA) proteins have been proved to play a critical role in determining viral infectivity and pandemic potential [125, 126]. Numerous studies used either computational methods or deep learning approaches to predict the mutation patterns of hemagglutinin [127, 128, 129]. [130] linked

the indicators of evolutionary changes in influenza based on changes in HA and NA to regional epidemic characteristics. Evolutionary metrics was introduced by analyzing amino acid substitutions in antigenically important regions of HA and modulated the transmission rate in the compartmental model [131]. While deriving metrics from HA and NA protein sequences can effectively capture many aspects of the virus's evolutionary dynamics, the computational approaches employed may overlook some critical features.

To include the full biological sequence data along with the influenza surveillance data, a deep learning framework was preferred. [132] reviewed state-of-the-art deep learning architectures applied to time series forecasting and highlighted that deep learning models were designed to capture intricate temporal patterns and nonlinear relationships that classical models often miss. It also mentioned that deep learning models may lack interpretability, which remained crucial for influenza forecasting. N-beats, a novel deep learning architecture, used fully connected layers arranged into deep stacks and employed both forward and backward residual connections [133]. The special structure can decompose the forecast into components and provide interpretability. Moreover, diffusion models, widely used in the generation of images, audio and text, were used to generate long-term time-series forecasts with improved stability [134]. A N-beats structure with diffusion models handling the multi-strain influenza time series data can generate stable long-term predictions, include influenza virus's HA and NA evolution, and provide interpretability.

6.2 Potential Methods and Data

6.2.1 Influenza surveillance data

Weekly surveillance data of influenza and weekly consultation rates of influenza-like illness reported by General Out-patient Clinics (GOPC) were obtained from the Hong Kong Department of Health's Centre for Health Protection. Severe influenza case, positive percentage for different influenza strain among all respiratory specimens and influenza-like illness(ILI) consultation rate in sentinel GOPC (per 1000 consultations) were used from 2014. Similar to previous research [29] the weekly ILI+ rate/severe influenza case+ was calculated by multiplying the ILI rate/total number of severe influenza case and the viral detection rate for each strain (A(H3N2), A(H1N1) and B) individually. Both ILI+ rate and severe influenza case+ were used for forecasting.

6.2.2 Human influenza HA and NA sequence

Human Influenza hemagglutinin (HA) and Neuraminidase (NA) A(H3N2), A(H1N1) and B sequences with full length were downloaded from NCBI Influenza Virus Resources. Duplicated data have been removed by keeping the oldest HA and NA sequences of each group. Sequences with unkonwn year were removed. Codon 17 to 345 for HA, codon 36 to 469 for NA were used as the valid positions. Dominant sequence of each month was determined as the sequence consisting of all the dominant amino acids in each codon in that month.

6.2.3 Weather and humidity data

Meteorological data, including daily temperature and relative humidity, were downloaded from the Hong Kong Observatory (HKO) website. Weekly maximum temperature, minimum temperature, mean temperature, and mean relative humidity were extracted from daily data.

6.2.4 N-beats Structure with long-term time series diffusion

An interpretable, multi-variate forecasting framework based on the N-BEATS architecture was proposed. It was structured as three additive stacks that provide influenza surveillance time series forecsating. The first stack is dedicated exclusively to the influenza surveillance signal through a multi-variate long-term diffusion module , capturing the temporal dynamics and interactions in influenza surveillance data. This stack is based on diffusion model to gradually adding small noise to influenza surveillance time series data and then training a neural network to learn the reverse process.

The second stack concats influenza surveillance data with weather and humidity data to account for exogenous influences on influenza dynamics. The third stack contains influenza protein sequence data, processed via a deep sequence encoder to extract biological signals indicative of viral evolution and strain dynamics and provide influenza forecasting. Each stack outputs both a forecast and a backcast (residual) that can be trained. The final forecast is aggregated by those three stacks, thus the long-term forecast can be interpreted as a combination of uncertainty and seasonality of influenza dynamic, environmental factors, and biological evolutionary effect.

Chapter 7

Conclusion

This study explored the impact of COVID-19, influenza, and air pollution on mortality with statistical methods and machine learning models, focusing on the COVID-19 pandemic period and post pandemic era.

Chapter 2 revealed two peaks of all-cause mortality during the COVID-19 pandemic in Peru with lower than usual median ages before Jan 2022. The excess mortality during this period mostly fell into diseases of the circulatory system, diseases of the respiratory system, general symptoms and sign, and COVID-19 compared to normal period. Meanwhile, the primary cause of death of most indirectly excess mortality was listed as diseases of the circulatory system and may be caused by the disruption in medical service during the pandemic.

Incubation periods of two variants of SARS-CoV-2 were estimated in Chapter 3 with two separate datasets collected in South Korea. The estimated mean incubation periods for 22 cases infected by Omicron BA.1 variants and 64 cases identified during Delta dominance were 3.5 days (95% CI: 2.5, 3.8) with 1.4 days (95% CI: 1.0, 1.5) SD and 6.5 days (95% CI: 5.3, 7.7) with 3.7 days (95% CI:

2.9, 4.6) respectively. A sensitivity analysis was performed by using either shorter or longer exposure bound, and a consistent estimation of incubation period was carried out. Based on the estimated distribution of incubation periods of both variants, a shorter incubation period of Omicron BA.1 may cause a relatively higher rate for Omicron BA.1 cases becoming new sources of infection. Hence, the Omicron BA.1 variants may have a lower period doubling time for epidemic curve. The Omicron BA.1 incubation period estimates of the 99-th percentile at 7.2 days further suggested that a 7-day quarantine combined with PCR testing could be sufficient to detect nearly 99% of Omicron BA.1 infections.

In Chapter 4, a new metric, daily exceedance concentration hours (DECH), reflected ambient $PM_{2.5}$ concentration was reviewed with Hong Kong mortality and air pollution data from 1999 to 2023. Despite the improvement of $PM_{2.5}$ concentration in Hong Kong over years, a fairly consistent and significant association between $PM_{2.5}$ concentration and all-cause mortality was found. Cause-specific mortalities, including circulatory diseases, respiratory diseases, mental and behavioral conditions, diseases of the nervous system and sense organs, and diseases of the skin and subcutaneous tissue, were examined and only mortality related to circulatory diseases reveals an association with the ambient $PM_{2.5}$ concentration. A sudden increase of coefficient in 2020 and a following plunge in 2021 and 2022 may be caused by the improvement of air pollution in 2020 and the surge of fatal cases of COVID-19 in 2021 and 2022.

Chapter 5 used both XGBoost and GAM model to estimate weekly mortality of different age groups in Hong Kong based on pre-pandemic data. The total excess mortality, the difference between estimated mortality and the observed death counts in 2020-2021, was calculated as well as disease burden attributable to in-

fluenza and air pollutants. In the first two years of the COVID-19 pandemic in Hong Kong, besides the 13 direct deaths related to COVID-19, 8762 (95% CI, 7503, 9993) and 12496 (95% CI, 11718, 13332) excess all-cause deaths were estimated, which were 12.6% and 8.5% of total deaths in 2020 and 2021 respectively. A notable shift in disease burden attributable to influenza and air pollutants was reported during the pandemic period compared to the previous five years and suggested when estimating the disease burden of the COVID-19 pandemic, both direct and indirect excess mortality should be considered.

The major limitation of this study is that it relied primarily on traditional statistical methods and machine learning models rather than deep learning models, despite their recent advance in time series prediction. Deep learning approaches usually have limited interpretability which is essential in understanding epidemic dynamics and accessing the impact of disease and air pollution on mortality. In Chapter 6, a future study on influenza forecasting using deep learning structure was discussed. The limited scope of available data (aggregated weekly instead of per-minute or per-second) further brings troubles on deploying deep learning models. This coarse temporal resolution and small dataset size needs more efforts to perform reliably training and validating. Yet, it is worth exploring new approaches in disease dynamics forecasting in order to generate more accurate predictions and estimates and provide insights in disease control and epidemic dynamics.

References

[1] Marco Ciotti et al. “The COVID-19 pandemic”. In: *Critical reviews in clinical laboratory sciences* 57.6 (2020), pp. 365–388.

[2] WHO. “Coronavirus disease (COVID-19) pandemic”. 2020.

[3] Malay Sarkar and Irappa Madabhavi. “COVID-19 mutations: An overview”. In: *World Journal of Methodology* 14.3 (2024), p. 89761.

[4] Christopher JL Murray. “COVID-19 will continue but the end of the pandemic is near”. In: *The Lancet* 399.10323 (2022), pp. 417–419.

[5] AV Raveendran, Rajeev Jayadevan, and S Sashidharan. “Long COVID: an overview”. In: *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 15.3 (2021), pp. 869–875.

[6] GJB Sousa et al. “Mortality and survival of COVID-19”. In: *Epidemiology & Infection* 148 (2020), e123.

[7] Esmaeil Mehraeen et al. “Predictors of mortality in patients with COVID-19–a systematic review”. In: *European journal of integrative medicine* 40 (2020), p. 101226.

[8] Haidong Wang et al. “Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21”. In: *The Lancet* 399.10334 (2022), pp. 1513–1536.

[9] ZS Venter, K Aunan, S Chowdhury, et al. “COVID-19 lockdowns cause global air pollution declines”. In: *Proceedings of the National Academy of Sciences* 117 (2020), pp. 18984–18990. DOI: 10.1073/pnas.2006853117.

[10] Susan Martin and Jonas Bergmann. “(Im) mobility in the age of COVID-19”. In: *International Migration Review* 55.3 (2021), pp. 660–687.

[11] Daniel A Solomon, Amy C Sherman, and Sanjat Kanjilal. “Influenza in the COVID-19 Era”. In: *Jama* 324.13 (2020), pp. 1342–1343.

[12] Todd W Lyons and Caroline G Kahane. “Changes in Respiratory Viral Testing Before and After the COVID-19 Pandemic”. In: *JAMA Network Open* 8.3 (2025), e250168–e250168.

[13] Shui Shan Lee, Cecile Viboud, and Eskild Petersen. “Understanding the rebound of influenza in the post COVID-19 pandemic period holds important clues for epidemiology and control”. In: *International Journal of Infectious Diseases* 122 (2022), pp. 1002–1004.

[14] Elliott T Wilkins. “Air pollution and the London fog of December, 1952”. In: *Journal of the Royal Sanitary Institute* 74.1 (1954), pp. 1–21.

[15] H Ross Anderson. “Air pollution and mortality: A history”. In: *Atmospheric Environment* 43.1 (2009), pp. 142–152.

[16] Markus Amann et al. “Reducing global air pollution: the scope for further policy interventions”. In: *Philosophical Transactions of the Royal Society A* 378.2183 (2020), p. 20190331.

[17] Michal Krzyzanowski and Aaron Cohen. “Update of WHO air quality guidelines”. In: *Air Quality, Atmosphere & Health* 1 (2008), pp. 7–13.

[18] World Health Organization. “Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide”. World Health Organization, 2006.

[19] World Health Organization. “WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide”. World Health Organization, 2021.

[20] Joel Schwartz. “Air pollution and daily mortality: a review and meta analysis”. In: *Environmental research* 64.1 (1994), pp. 36–52.

[21] Shui He, Sati Mazumdar, and Vincent C Arena. “A comparative study of the use of GAM and GLM in air pollution research”. In: *Environmetrics: The official journal of the International Environmetrics Society* 17.1 (2006), pp. 81–93.

[22] Roger D Peng et al. “Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients”. In: *Jama* 299.18 (2008), pp. 2172–2179.

[23] H. Lin et al. “Using daily excessive concentration hours to explore the short-term mortality effects of ambient PM 2.5 in Hong Kong”. In: *Environmental Pollution* 229 (2017), pp. 896–901. DOI: 10.1016/j.envpol.2017.07.060.

[24] Fang Guo et al. “Ozone as an environmental driver of influenza”. In: *Nature communications* 15.1 (2024), p. 3763.

[25] World Health Organization et al. “Influenza (seasonal) fact sheet”. In: *World Health Organization, Geneva, Switzerland. https://www. who. int/news-room/fact-sheets/detail/influenza-(seasonal)* (2025).

[26] Arnold S Monto and Keiji Fukuda. “Lessons from influenza pandemics of the last 100 years”. In: *Clinical Infectious Diseases* 70.5 (2020), pp. 951–957.

[27] Yu-Chia Hsieh et al. “Influenza pandemics: past, present and future”. In: *Journal of the Formosan Medical Association* 105.1 (2006), pp. 1–6.

[28] Aubree Gordon and Arthur Reingold. “The burden of influenza: a complex problem”. In: *Current epidemiology reports* 5 (2018), pp. 1–9.

[29] Wan Yang et al. “Forecasting influenza epidemics in Hong Kong”. In: *PLoS computational biology* 11.7 (2015), e1004383.

[30] L Yang, K Chan, B Cowling, et al. “Excess mortality associated with the 2009 pandemic of influenza A (H1N1) in Hong Kong”. In: *Epidemiology & Infection* 140 (2012), pp. 1542–1550. DOI: 10 . 1017 / S0950268811002238.

[31] A Danielle Iuliano et al. “Estimates of global seasonal influenza-associated respiratory mortality: a modelling study”. In: *The Lancet* 391.10127 (2018), pp. 1285–1300.

[32] Nicholas G Reich et al. “A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States”. In: *Proceedings of the National Academy of Sciences* 116.8 (2019), pp. 3146–3154.

[33] Seungwon Jung et al. “Self-attention-based deep learning network for regional influenza forecasting”. In: *IEEE Journal of Biomedical and Health Informatics* 26.2 (2021), pp. 922–933.

[34] Qinneng Xu et al. “Forecasting influenza in Hong Kong with Google search queries and statistical model fusion”. In: *PloS one* 12.5 (2017), e0176690.

[35] Liuyang Yang et al. “Deep-learning model for influenza prediction from multisource heterogeneous data in a megacity: model development and evaluation”. In: *Journal of Medical Internet Research* 25 (2023), e44238.

[36] Percy Herrera-Añazco et al. “Some lessons that Peru did not learn before the second wave of COVID-19”. In: *The International Journal of Health Planning and Management* (2021).

[37] Thomas Beaney et al. “Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide?” In: *Journal of the Royal Society of Medicine* 113.9 (2020), pp. 329–334.

[38] Luke Taylor. “Covid-19: Why Peru suffers from one of the highest excess death rates in the world”. In: *bmj* 372 (2021).

[39] Cesar Munayco et al. “Risk of death by age and gender from CoVID-19 in Peru, March-May, 2020”. In: *Aging (Albany NY)* 12.14 (2020), p. 13869.

[40] Zelalem G Dessie and Temesgen Zewotir. “Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients”. In: *BMC infectious diseases* 21.1 (2021), pp. 1–28.

[41] B Lutz and M E. “During the Omicron Wave, Death Rates Soared for Older People”. In: *The New York Times* (2022).

[42] Yuelong Shu and John McCauley. “GISAID: Global initiative on sharing all influenza data–from vision to reality”. In: *Eurosurveillance* 22.13 (2017), p. 30494.

[43] Ilze Gobiņa et al. “Excess mortality associated with the COVID-19 pandemic in Latvia: a population-level analysis of all-cause and noncommunicable disease deaths in 2020”. In: *BMC Public Health* 22.1 (2022), pp. 1–12.

[44] World Health Organization. “COVID-19 Significantly Impacts Health Services for Noncommunicable Diseases”. https://www.who.int/news-room/detail/01-06-2020-covid-19-significantly-impacts-health-services-for-noncommunicable-diseases. 2020.

[45] Diego Chambergo-Michilot et al. “Trends and geographical variation in mortality from coronary disease in Peru”. In: *PloS one* 17.9 (2022), e0273949.

[46] Léa Sempé et al. “Estimation of all-cause excess mortality by age-specific mortality patterns for countries with incomplete vital statistics: a population-based study of the case of Peru during the first wave of

the COVID-19 pandemic". In: *The Lancet Regional Health-Americas* 2 (2021), p. 100039.

[47] Mohammad Madjid et al. "Potential effects of coronaviruses on the cardiovascular system: a review". In: *JAMA cardiology* 5.7 (2020), pp. 831–840.

[48] Thirumalaisamy P Velavan and Christian G Meyer. "The COVID-19 epidemic". In: *Tropical medicine & international health* 25.3 (2020), p. 278.

[49] Moritz UG Kraemer et al. "Monitoring key epidemiological parameters of SARS-CoV-2 transmission". In: *Nature medicine* 27.11 (2021), pp. 1854–1855.

[50] Yuto Homma et al. "The incubation period of the SARS-CoV-2 B1. 1.7 variant is shorter than that of other strains". In: *Journal of Infection* 83.2 (2021), e15–e17.

[51] Kenrad E Nelson and Carolyn Masters Williams. "Infectious disease epidemiology: theory and practice". Jones & Bartlett Publishers, 2014.

[52] Hongjie Xin et al. "The incubation period distribution of coronavirus disease 2019: a systematic review and meta-analysis". In: *Clinical Infectious Diseases* 73.12 (2021), pp. 2344–2352.

[53] Stephen A Lauer et al. "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application". In: *Annals of internal medicine* 172.9 (2020), pp. 577–582.

[54] Hongjie Xin et al. "Estimating the latent period of coronavirus disease 2019 (COVID-19)". In: *Clinical Infectious Diseases* (2021).

[55] Conor McAloon et al. “Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research”. In: *BMJ open* 10.8 (2020), e039652.

[56] Sang-Won Ryu et al. “Serial interval and transmission dynamics during SARS-CoV-2 delta variant predominance, South Korea”. In: *Emerging infectious diseases* 28.2 (2022), p. 407.

[57] Donghyun Kim et al. “Estimation of Serial Interval and Reproduction Number to Quantify the Transmissibility of SARS-CoV-2 Omicron Variant in South Korea”. In: *Viruses* 14.3 (2022), p. 533.

[58] Shi Zhao et al. “Estimating the serial interval of the novel coronavirus disease (COVID-19): a statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020”. In: *Frontiers in Physics* 8 (2020), p. 347.

[59] Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. “Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020”. In: *Eurosurveillance* 25.5 (2020), p. 2000062.

[60] Sang Woo Park et al. “Forward-looking serial intervals correctly link epidemic growth to reproduction numbers”. In: *Proceedings of the National Academy of Sciences* 118.2 (2021).

[61] Yuk Ching Lau et al. “Joint Estimation of Generation Time and Incubation Period for Coronavirus Disease 2019”. In: *The Journal of infectious diseases* 224.10 (2021), pp. 1664–1671.

[62] William S Hart et al. “Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis”. In: *The Lancet Infectious Diseases* (2022).

[63] Min Kang et al. “Transmission dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant infections in Guangdong, China, May to June 2021”. In: *Eurosurveillance* 27.10 (2022), p. 2100815.

[64] Xi He et al. “Temporal dynamics in viral shedding and transmissibility of COVID-19”. In: *Nature medicine* 26.5 (2020), pp. 672–675.

[65] Åke Svensson. “A note on generation times in epidemic models”. In: *Mathematical biosciences* 208.1 (2007), pp. 300–311.

[66] Dirk Eggink et al. “Increased risk of infection with SARS-CoV-2 Omicron BA. 1 compared with Delta in vaccinated and previously infected individuals, the Netherlands, 22 November 2021 to 19 January 2022”. In: *Eurosurveillance* 27.4 (2022), p. 2101196.

[67] J Yu et al. “Neutralization of the SARS-CoV-2 Omicron BA. 1 and BA. 2 Variants”. In: *New England Journal of Medicine* (2022).

[68] Michael A Johansson et al. “Reducing travel-related SARS-CoV-2 transmission with layered mitigation measures: symptom monitoring, quarantine, and testing”. In: *BMC medicine* 19.1 (2021), pp. 1–13.

[69] Johannes Henssler et al. “Mental health effects of infection containment strategies: quarantine and isolation—a systematic review and meta-analysis”. In: *European archives of psychiatry and clinical neuroscience* 271.2 (2021), pp. 223–234.

[70] Anhui Tang et al. “Detection of novel coronavirus by RT-PCR in stool specimen from asymptomatic child, China”. In: *Emerging infectious diseases* 26.6 (2020), p. 1337.

[71] Thomas A Treibel et al. “COVID-19: PCR screening of asymptomatic health-care workers at London hospital”. In: *The Lancet* 395.10237 (2020), pp. 1608–1610.

[72] M. Franklin, A. Zeka, and J. Schwartz. “Association between PM 2.5 and all-cause and specific-cause mortality in 27 US communities”. In: *Journal of exposure science & environmental epidemiology* 17.3 (2007), pp. 279–287. DOI: 10.1038/sj.jes.7500530.

[73] J. S. Apte et al. “Addressing global mortality from ambient PM2.5”. In: *Environmental science & technology* 49.13 (2015), pp. 8057–8066. DOI: 10.1021/acs.est.5b01236.

[74] R. E. Dales and S. Cakmak. “Does mental health status influence susceptibility to the physiologic effects of air pollution? A population based study of Canadian children”. In: *PLoS One* 11.12 (2016), e0168931. DOI: 10.1371/journal.pone.0168931.

[75] S. Genc et al. “The adverse effects of air pollution on the nervous system”. In: *Journal of toxicology* 2012 (2012). DOI: 10.1155/2012/782462.

[76] K. E. Kim, D. Cho, and H. J. Park. “Air pollution and skin diseases: Adverse effects of airborne particulate matter on various skin diseases”. In: *Life sciences* 152 (2016), pp. 126–134. DOI: 10.1016/j.lfs.2016.03.039.

[77] Yanwen Liu et al. “Change in disease burden associated with influenza and air pollutants during the COVID-19 pandemic in Hong Kong”. In: *Digital Health* 10 (2024), p. 20552076241261892.

[78] K. Chen et al. “Air pollution reduction and mortality benefit during the COVID-19 outbreak in China”. In: *The Lancet Planetary Health* 4.6 (2020), e210–e212.

[79] H. C. Ho et al. “Spatiotemporal influence of temperature, air quality, and urban environment on cause-specific mortality during hazy days”. In: *Environment international* 112 (2018), pp. 10–22. DOI: 10.1016/j.envint.2017.12.001.

[80] L. Calderón-Garcidueñas et al. “Megacities air pollution problems: Mexico City Metropolitan Area critical issues on the central nervous system pediatric impact”. In: *Environmental research* 137 (2015), pp. 157–169. DOI: 10.1016/j.envres.2014.12.012.

[81] H. Lin et al. “Hourly peak concentration measuring the PM 2.5 -mortality association: Results from six cities in the Pearl River Delta study”. In: *Atmospheric Environment* 161 (2017), pp. 27–33. DOI: 10.1016/j.atmosenv.2017.04.015.

[82] H. Qiu et al. “Effects of coarse particulate matter on emergency hospital admissions for respiratory diseases: a time series analysis in Hong Kong”. In: *Environmental Health Perspectives* 120 (2012), pp. 572–576.

[83] Center for Health Protection. “Flu Express”. 11 (14). 2014.

[84] H. Lin et al. “Daily exceedance concentration hours: A novel indicator to measure acute cardiovascular effects of PM2.5 in six Chinese subtropical cities”. In: *Environment International* 111 (2017), pp. 117–123. DOI: 10.1016/j.envint.2017.11.022.

[85] L. Tian et al. “Ambient carbon monoxide associated with reduced risk of hospital admissions for respiratory tract infections”. In: *American Journal of Respiratory and Critical Care Medicine* 188.10 (2013), pp. 1240–1245.

[86] H. C. Ho et al. “The associations between social, built and geophysical environment and age-specific dementia mortality among older adults in a high-density Asian city”. In: *International Journal of Health Geographics* 19.1 (2020), pp. 1–13. DOI: 10.1186/s12942-020-00252-y.

[87] CK Lai, RW Ng, MC Wong, et al. “Epidemiological characteristics of the first 100 cases of coronavirus disease 2019 (COVID-19) in Hong Kong Special Administrative Region, China, a city with a stringent containment policy”. In: *International journal of epidemiology* 49 (2020), pp. 1096–1105. DOI: 10.1093/ije/dyaa106.

[88] MC Wong, RW Ng, KC Chong, et al. “Stringent containment measures without complete city lockdown to achieve low incidence and mortality across two waves of COVID-19 in Hong Kong”. In: *BMJ global health* 5 (2020), e003573.

[89] WC Koh, MF Alikhan, D Koh, et al. “Containing COVID-19: implementation of early and moderately stringent social distancing measures can prevent the need for large-scale lockdowns”. In: *Annals of global health* 86 (2020).

[90] SAR HK. “Government further tightens social distancing measures”. https://www.info.gov.hk/gia/general/202007/27/P2020072700650.htm?fontSize=1. 2020.

[91] Health CfHPotDo. “Archive of Statistics on Provisional Analysis on Reported Death Cases”. https://www.coronavirus.gov.hk/eng/death_analysis.html. 2022, 2023.

[92] N Jones. “How COVID-19 is changing the cold and flu season”. In: *Nature* 588 (2020), pp. 388–390.

[93] KH Chan, P-w Lee, CY Chan, et al. “Monitoring respiratory infections in covid-19 epidemics”. In: *BMJ* 369 (2020), p. m1628. DOI: 10.1136/bmj.m1628.

[94] L Yang, X Wang, K Chan, et al. “Hospitalisation associated with the 2009 H1N1 pandemic and seasonal influenza in Hong Kong, 2005 to 2010”. In: *Eurosurveillance* 17 (2012).

[95] Tiffany A Walker et al. “Risk of severe influenza among adults with chronic medical conditions”. In: *The Journal of infectious diseases* 221.2 (2020), pp. 183–190.

[96] C Signorelli, T Scognamiglio, and A Odone. “COVID-19 in Italy: impact of containment measures and prevalence estimates of infection in the general population”. In: *Acta Bio Medica: Atenei Parmensis* 91 (2020), p. 175. DOI: 10.23750/abm.v91i3-S.9511.

[97] H Ritchie, E Mathieu, L Rodés-Guirao, et al. “Coronavirus pandemic (COVID-19)”. https://ourworldindata.org/coronavirus. 2020.

[98] Lauren M Rossen et al. “Excess all-cause mortality in the USA and Europe during the COVID-19 pandemic, 2020 and 2021”. In: *Scientific reports* 12.1 (2022), p. 18559.

[99] P Wu, E Goldstein, LM Ho, et al. “Excess mortality associated with influenza A and B virus in Hong Kong, 1998–2009”. In: *The Journal of infectious diseases* 206 (2012), pp. 1862–1871. DOI: 10.1093/infdis/jis628.

[100] X-L Wang, L Yang, K-P Chan, et al. “Model selection in time series studies of influenza-associated mortality”. In: *PLoS One* 7 (2012), e39423. DOI: 10.1371/journal.pone.0039423.

[101] X Yu, C Wang, T Chen, et al. “Excess pneumonia and influenza mortality attributable to seasonal influenza in subtropical Shanghai, China”. In: *BMC Infectious Diseases* 17 (2017), pp. 1–9. DOI: 10.1186/s12879-017-2863-1.

[102] VS Desdhanty and Z Rustam. “Liver cancer classification using random forest and extreme gradient boosting (xgboost) with genetic algorithm as feature selection”. In: *2021 International Conference on Decision Aid Sciences and Application (DASA)*. IEEE. 2021, pp. 716–719.

[103] S Dhillon, C Bansal, and B Sidhu. “Machine Learning Based Approach Using XGboost for Heart Stroke Prediction”. In: *International Conference on Emerging Technologies: AI, IoT, and CPS for Science & Technology Applications*. 2021, pp. 06–07.

[104] J Ding, M Liu, Z Ma, et al. “Spatial and temporal trends in the mortality burden of ozone pollution in China: 2005-2017”. In: *ISEE Conference Abstracts*. 2020.

[105] JY Wong, P Wu, H Nishiura, et al. “Infection fatality risk of the pandemic A (H1N1) 2009 virus in Hong Kong”. In: *American journal of epidemiology* 177 (2013), pp. 834–840. DOI: 10.1093/aje/kws314.

[106] DE Hilt and DW Seegrist. “Ridge, a computer program for calculating ridge regression estimates”. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 1977.

[107] H Xin, P Wu, JY Wong, et al. “Hospitalizations and mortality during the first year of the COVID-19 pandemic in Hong Kong, China: An observational study”. In: *The Lancet Regional Health–Western Pacific* 30 (2023).

[108] NC Chiu, H Chi, YL Tai, et al. “Impact of Wearing Masks, Hand Hygiene, and Social Distancing on Influenza, Enterovirus, and All-Cause Pneumonia During the Coronavirus Pandemic: Retrospective National Epidemiological Surveillance Study”. In: *J Med Internet Res* 22 (2020), e21257. DOI: 10.2196/21257.

[109] Eric J Chow, Timothy M Uyeki, and Helen Y Chu. “The effects of the COVID-19 pandemic on community respiratory virus activity”. In: *Nature Reviews Microbiology* 21.3 (2023), pp. 195–210.

[110] SAR HK. “Statistics on seasonal influenza vaccination”. https://www.info.gov.hk/gia/general/202011/18/P2020111800593.htm?fontSize=1. 2020.

[111] Z Xu, R Cao, X Hu, et al. “The improvement of air quality and associated mortality during the COVID-19 lockdown in one megacity of China: an empirical strategy”. In: *International Journal of Environmental Research and Public Health* 18 (2021), p. 8702. DOI: 10.3390/ijerph18168702.

[112] ZS Venter, K Aunan, S Chowdhury, et al. “Air pollution declines during COVID-19 lockdowns mitigate the global health burden”. In: *Environmental research* 192 (2021), p. 110403. DOI: 10.1016/j.envres.2020.110403.

[113] B Davies, BL Parkes, J Bennett, et al. “Community factors and excess mortality in first wave of the COVID-19 pandemic in England”. In: *Nature communications* 12 (2021), p. 3755. DOI: 10.1038/s41467-021-23935-x.

[114] RG Ryan, JD Silver, and R Schofield. “Air quality and health impact of 2019–20 Black Summer megafires and COVID-19 lockdown in Melbourne and Sydney, Australia”. In: *Environmental Pollution* 274 (2021), p. 116498. DOI: 10.1016/j.envpol.2021.116498.

[115] PA Kowalski, M Szwagrzyk, J Kielpinska, et al. “Numerical analysis of factors, pace and intensity of the corona virus (COVID-19) epidemic in Poland”. In: *Ecological informatics* 63 (2021), p. 101284. DOI: 10.1016/j.ecoinf.2021.101284.

[116] L Becchetti, G Beccari, G Conzo, et al. “Air quality and COVID-19 adverse outcomes: Divergent views and experimental findings”. In: *Environmental research* 193 (2021), p. 110556. DOI: 10.1016/j.envres.2020.110556.

[117] P Yin, M Brauer, AJ Cohen, et al. “The effect of air pollution on deaths, disease burden, and life expectancy across China and its provinces, 1990–2017: an analysis for the Global Burden of Disease Study 2017”. In: *Lancet Planet Health* 4 (2020), E386–E398. DOI: 10.1016/S2542-5196(20)30161-3.

[118] XC Lu, T Yao, JCH Fung, et al. “Estimation of health and economic costs of air pollution over the Pearl River Delta region in China”. In: *Sci Total Environ* 566 (2016), pp. 134–143. DOI: 10.1016/j.scitotenv.2016.05.060.

[119] Marc Lipsitch et al. “Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1”. In: *Biosecurity and bioterrorism: biodefense strategy, practice, and science* 9.2 (2011), pp. 89–115.

[120] U.S. Centers for Disease Control and Prevention. “FluSight Challenge”. Accessed: 13 May 2025. 2025. URL: https://www.cdc.gov/flu/weekly/.

[121] Wan Yang, Eric HY Lau, and Benjamin J Cowling. “Dynamic interactions of influenza viruses in Hong Kong during 1998-2018”. In: *PLoS computational biology* 16.6 (2020), e1007989.

[122] Yunhao Liu et al. “Forecasting influenza epidemics in Hong Kong using Google search queries data: A new integrated approach”. In: *Expert Systems with Applications* 185 (2021), p. 115604.

[123] Nan Zhang et al. “Effects of human behavior changes during the coronavirus disease 2019 (COVID-19) pandemic on influenza spread in Hong Kong”. In: *Clinical Infectious Diseases* 73.5 (2021), e1142–e1150.

[124] Martha I Nelson and Edward C Holmes. “The evolution of epidemic influenza”. In: *Nature reviews genetics* 8.3 (2007), pp. 196–205.

[125] Charles J Russell, Meng Hu, and Faten A Okda. “Influenza hemagglutinin protein stability, activation, and pandemic risk”. In: *Trends in microbiology* 26.10 (2018), pp. 841–853.

[126] Ravendra P Chauhan and Michelle L Gordon. “An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates”. In: *Virus genes* 58.4 (2022), pp. 255–269.

[127] Yuan-Ling Xia et al. “A deep learning approach for predicting antigenic variation of influenza A H3N2”. In: *Computational and mathematical methods in medicine* 2021.1 (2021), p. 9997669.

[128] Marta Łuksza and Michael Lässig. “A predictive fitness model for influenza”. In: *Nature* 507.7490 (2014), pp. 57–61.

[129] Jiankui He and Michael W Deem. “Low-dimensional clustering detects incipient dominant influenza strain clusters”. In: *Protein Engineering, Design & Selection* 23.12 (2010), pp. 935–946.

[130] Amanda C Perofsky et al. “Antigenic drift and subtype interference shape A (H3N2) epidemic dynamics in the United States”. In: *Elife* 13 (2024), RP91849.

[131] Xiangjun Du et al. “Evolution-informed forecasting of seasonal influenza A (H3N2)”. In: *Science translational medicine* 9.413 (2017), eaan5325.

[132] Angelo Casolaro et al. “Deep learning for time series forecasting: Advances and open problems”. In: *Information* 14.11 (2023), p. 598.

[133] Boris N Oreshkin et al. “N-BEATS: Neural basis expansion analysis for interpretable time series forecasting”. In: *arXiv preprint arXiv:1905.10437* (2019).

[134] Lifeng Shen and James Kwok. “Non-autoregressive conditional diffusion models for time series prediction”. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 31016–31029.

Appendices

9.1 Trends in the effects of ambient PM2.5 concentration on mortality risk in Hong Kong, China

Table 9.1: DECH Coefficient Data pertaining to Figure 4.2

Window End Year	All (Total)	All (Lower)	All (Upper)	Circ (Total)	Circ (Lower)	Circ (Upper)	Resp (Total)	Resp (Lower)	Resp (Upper)
2011	1.70E-05	7.92E-06	2.60E-05	2.79E-05	1.13E-05	4.44E-05	2.08E-05	2.81E-07	4.13E-05
2012	1.49E-05	5.81E-06	2.40E-05	2.79E-05	1.09E-05	4.50E-05	1.50E-05	-5.22E-06	3.52E-05
2013	1.66E-05	7.69E-06	2.55E-05	3.30E-05	1.62E-05	4.98E-05	1.48E-05	-5.18E-06	3.47E-05
2014	1.63E-05	7.24E-06	2.54E-05	3.44E-05	1.73E-05	5.16E-05	1.31E-05	-7.02E-06	3.33E-05
2015	1.66E-05	7.42E-06	2.57E-05	3.20E-05	1.46E-05	4.94E-05	1.53E-05	-4.99E-06	3.56E-05
2016	1.50E-05	5.72E-06	2.43E-05	3.15E-05	1.34E-05	4.95E-05	1.44E-05	-5.91E-06	3.46E-05
2017	1.37E-05	3.92E-06	2.35E-05	4.08E-05	2.17E-05	5.99E-05	8.33E-06	-1.27E-05	2.93E-05
2018	1.64E-05	6.16E-06	2.66E-05	4.26E-05	2.23E-05	6.29E-05	1.13E-05	-1.03E-05	3.29E-05
2019	1.81E-05	7.38E-06	2.89E-05	4.44E-05	2.29E-05	6.59E-05	1.43E-05	-8.25E-06	3.69E-05
2020	3.00E-05	1.84E-05	4.16E-05	5.73E-05	3.38E-05	8.08E-05	2.85E-05	4.61E-06	5.24E-05
2021	1.82E-05	5.92E-06	3.06E-05	3.88E-05	1.34E-05	6.43E-05	5.42E-06	-1.95E-05	3.03E-05
2022	2.81E-06	-1.72E-05	2.28E-05	2.96E-05	2.06E-06	5.72E-05	-2.79E-05	-7.27E-05	1.68E-05
2023	6.68E-06	-1.48E-05	2.81E-05	3.53E-05	6.24E-06	6.44E-05	-2.75E-05	-7.84E-05	2.33E-05

Table 9.2: DECH IQR values for each window ending in given years during a 13-year sliding window.

Window End Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
DECH IQR	608	605	613	611	616	585	538	499	454	415	370	338	307

Table 9.3: DECH coefficient data pertaining to Figure 4.3

Window End Year	All	All (lower)	All (upper)	Circ	Circ (lower)	Circ (upper)	Resp	Resp (lower)	Resp (upper)
2003	2.31E-05	6.54E-06	3.96E-05	2.79E-05	-1.22E-06	5.69E-05	3.59E-05	-4.22E-06	7.61E-05
2004	2.68E-05	1.20E-05	4.17E-05	1.90E-05	-8.47E-06	4.64E-05	4.98E-05	1.40E-05	8.56E-05
2005	2.56E-05	1.15E-05	3.98E-05	2.86E-05	3.14E-06	5.40E-05	3.19E-05	-1.24E-06	6.50E-05
2006	2.18E-05	8.12E-06	3.54E-05	2.47E-05	-8.72E-09	4.94E-05	3.51E-05	3.55E-06	6.67E-05
2007	6.61E-06	-6.33E-06	1.96E-05	1.33E-05	-1.06E-05	3.72E-05	-5.89E-06	-3.54E-05	2.37E-05
2008	1.43E-05	1.58E-06	2.70E-05	2.78E-05	3.80E-06	5.17E-05	1.98E-05	-8.38E-06	4.79E-05
2009	1.01E-05	-3.16E-06	2.34E-05	3.92E-05	1.45E-05	6.40E-05	5.26E-06	-2.37E-05	3.43E-05
2010	8.03E-06	-5.89E-06	2.19E-05	3.68E-05	1.00E-05	6.36E-05	2.08E-06	-2.75E-05	3.17E-05
2011	9.69E-06	-5.14E-06	2.45E-05	3.36E-05	5.45E-06	6.17E-05	9.64E-07	-3.05E-05	3.24E-05
2012	1.60E-05	2.47E-07	3.18E-05	2.76E-05	-2.85E-06	5.80E-05	2.08E-05	-1.25E-05	5.40E-05
2013	8.40E-06	-7.30E-06	2.41E-05	2.06E-05	-1.09E-05	5.20E-05	-1.13E-06	-3.47E-05	3.24E-05
2014	1.04E-05	-6.56E-06	2.74E-05	4.20E-05	7.21E-06	7.67E-05	-1.26E-05	-4.83E-05	2.31E-05
2015	1.83E-05	5.65E-07	3.60E-05	2.94E-05	-6.90E-06	6.58E-05	8.48E-06	-2.95E-05	4.64E-05
2016	2.21E-05	3.44E-06	4.07E-05	3.90E-05	-4.05E-07	7.84E-05	2.03E-05	-1.83E-05	5.89E-05
2017	2.96E-05	1.00E-05	4.92E-05	7.33E-05	3.26E-05	0.00114	2.65E-05	-1.36E-05	6.66E-05
2018	3.68E-05	1.51E-05	5.85E-05	6.04E-05	1.41E-05	0.00107	2.52E-05	-1.81E-05	6.86E-05
2019	3.28E-05	1.07E-05	5.48E-05	3.58E-05	-1.31E-05	8.48E-05	3.10E-05	-1.31E-05	7.50E-05
2020	2.71E-05	1.56E-06	5.27E-05	6.90E-05	1.27E-05	0.00125	4.85E-06	-4.52E-05	5.49E-05
2021	2.72E-05	1.97E-07	5.42E-05	8.44E-05	2.55E-05	0.00143	-9.38E-06	-6.18E-05	4.31E-05
2022	2.41E-05	-2.71E-05	7.54E-05	6.06E-05	-4.66E-06	0.00126	3.14E-05	-7.20E-05	0.00135
2023	-3.06E-06	-6.63E-05	6.02E-05	5.31E-05	-1.85E-05	0.00125	-3.92E-05	-0.00016972	9.14E-05

Table 9.4: Relative effect percentages (%) from the DECH coefficients in Figure 4.4

Window End Year	All	All (lower)	All (upper)	Circ	Circ (lower)	Circ (upper)	Resp	Resp (lower)	Resp (upper)
2003	1.25	0.35	2.14	1.51	-0.07	3.08	1.94	-0.23	4.11
2004	1.57	0.7	2.45	1.11	-0.5	2.72	2.92	0.82	5.02
2005	1.64	0.73	2.54	1.82	0.2	3.44	2.03	-0.08	4.15
2006	1.47	0.55	2.39	1.67	0	3.34	2.37	0.24	4.5
2007	0.48	-0.46	1.41	0.96	-0.76	2.69	-0.42	-2.56	1.71
2008	1.04	0.12	1.97	2.03	0.28	3.78	1.44	-0.61	3.5
2009	0.68	-0.21	1.56	2.62	0.97	4.26	0.35	-1.58	2.28
2010	0.51	-0.37	1.39	2.33	0.64	4.02	0.13	-1.74	2.01
2011	0.59	-0.31	1.49	2.04	0.33	3.75	0.06	-1.85	1.97
2012	0.86	0.01	1.71	1.48	-0.15	3.12	1.12	-0.67	2.91
2013	0.44	-0.38	1.25	1.07	-0.57	2.71	-0.06	-1.81	1.69
2014	0.54	-0.34	1.41	2.16	0.37	3.96	-0.65	-2.49	1.19
2015	0.91	0.03	1.79	1.46	-0.34	3.27	0.42	-1.47	2.31
2016	0.89	0.14	1.65	1.58	-0.02	3.17	0.82	-0.74	2.39
2017	1.08	0.37	1.8	2.68	1.19	4.16	0.97	-0.5	2.43
2018	1.13	0.47	1.8	1.86	0.43	3.28	0.78	-0.56	2.11
2019	0.88	0.29	1.48	0.97	-0.35	2.28	0.83	-0.35	2.02
2020	0.62	0.04	1.21	1.59	0.29	2.88	0.11	-1.04	1.26
2021	0.58	0	1.15	1.79	0.54	3.04	-0.2	-1.31	0.91
2022	0.43	-0.48	1.33	1.07	-0.08	2.22	0.55	-1.27	2.38
2023	-0.05	-0.99	0.9	0.8	-0.28	1.87	-0.59	-2.55	1.37

Table 9.5: DECH IQR values for windows ending in each given year during a 5-year sliding window.

Window End Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
DECH IQR	541	587	638	675	722	730	667	633	608	538	520	516	498	405
Window End Year	2017	2018	2019	2020	2021	2022	2023							
DECH IQR	365	308	270	230	212	176	150							

Table 9.6: DECH Coefficient data pertaining to Figure 4.5

Window End Year	Ment	Ment (lower)	Ment (upper)	Nerv	Nerv (lower)	Nerv (upper)	Skin	Skin (lower)	Skin (upper)
2003	-5.40E-05	-0.000239986	0.0001322	2.02E-05	-0.000133645	0.0001741	-0.000167418	-0.000429483	9.46E-05
2004	-1.71E-05	-0.000162917	0.0001291	2.18E-05	-0.000120625	0.0001641	-7.55E-05	-0.000277094	0.0001261
2005	4.71E-06	-0.000122375	0.0001331	4.93E-05	-8.13E-05	0.0001801	2.17E-05	-0.000135762	0.0001791
2006	2.51E-05	-9.89E-05	0.0001491	3.59E-05	-9.23E-05	0.0001641	4.92E-05	-9.51E-05	0.0001931
2007	5.03E-06	-0.000117436	0.0001271	2.03E-05	-0.000105177	0.0001461	5.62E-05	-8.00E-05	0.0001921
2008	-7.79E-06	-0.000128378	0.0001131	6.21E-05	-6.28E-05	0.0001871	5.76E-05	-7.72E-05	0.0002201
2009	-4.36E-05	-0.000164874	7.76E-05	4.42E-05	-9.33E-05	0.0001821	7.75E-05	-6.51E-05	0.0002221
2010	-2.76E-05	-0.000143469	8.84E-05	3.95E-05	-0.000102886	0.0001821	3.39E-05	-0.000125975	0.0001941
2011	-4.34E-05	-0.000155684	6.89E-05	6.48E-05	-8.37E-05	0.0002131	-1.40E-05	-0.000191657	0.0001641
2012	-1.91E-05	-0.000128589	9.04E-05	-1.16E-05	-0.000177506	0.0001541	-5.98E-05	-0.000266508	0.0001471
2013	1.36E-05	-8.83E-05	0.0001161	2.06E-05	-0.000145542	0.0001871	-3.17E-05	-0.000255054	0.0001921
2014	7.39E-05	-2.89E-05	0.0001771	-2.86E-05	-0.000201084	0.0001441	-2.56E-05	-0.000268754	0.0002171
2015	1.98E-05	-8.12E-05	0.0001211	-6.86E-05	-0.000246831	0.0001111	-0.000102631	-0.000353423	0.0001481
2016	4.71E-05	-5.48E-05	0.0001491	9.80E-07	-0.000182638	0.0001851	-9.68E-05	-0.000359713	0.0001661
2017	6.15E-05	-3.95E-05	0.0001631	6.88E-05	-0.000119819	0.0002571	-0.000107133	-0.000386796	0.0001731
2018	8.67E-05	-2.07E-05	0.0001941	1.63E-05	-0.000193479	0.0002261	-0.000174023	-0.000473674	0.0001261
2019	6.42E-05	-4.52E-05	0.0001741	7.31E-05	-0.000138862	0.0002851	-5.91E-05	-0.000375077	0.0002571
2020	7.27E-05	-4.85E-05	0.0001941	3.57E-05	-0.000206688	0.0002781	-6.94E-07	-0.000353642	0.0003521
2021	8.19E-05	-4.40E-05	0.0002081	1.90E-06	-0.000246685	0.0002501	-7.66E-05	-0.000457588	0.0003041
2022	2.17E-05	-0.000116811	0.0001601	-8.19E-05	-0.000339122	0.0001751	-3.20E-05	-0.000427783	0.0003641
2023	-4.74E-05	-0.000201458	0.0001071	-3.38E-05	-0.000303651	0.0002361	-1.35E-05	-0.000453781	0.0004271

9.2 Estimating the incubation period of SARS-CoV-2 Omicron BA.1 variant in comparison with that during the Delta variant dominance in South Korea

Table 9.7: Summary of incubation period estimates (unit: day) for cases infected by Omicron BA.1 variants and cases during Delta predominance period in South Korea with different version of exposure bound.

Type of SARS-CoV-2 strain	sample size	mean of exposure distribution (days)	estimate (95%CI)		
			mean	median	95-th percentile (SD)
Omicron BA.1	$n = 22$	2.8	$3.5(2.5, 3.8)$	$3.3(2.4, 3.6)$	$5.8(4.3, 6.6)$ (1.4 $\pm$ 1.0, 1.5)
		4.6	$3.5(2.6, 3.9)$	$3.3(2.4, 3.7)$	$6.0(4.4, 6.6)$ (1.3 $\pm$ 1.0, 1.5)
those during Delta dominance[$]	$n = 64$	2.8	$6.0(4.7, 7.1)$	$5.1(3.7, 6.4)$	$13.6(11.0, 16.4)$ (3.9 $\pm$ 3.0, 5.2)
		4.6	$7.0(5.7, 8.2)$	$6.4(5.1, 7.7)$	$13.8(11.4, 16.0)$ (3.6 $\pm$ 2.8, 4.5)

[$] These cases were collected in June 2021 when the Delta variants were dominant at a prevalence of 68.3% in South Korea according to GISAID [42].

Table 9.8: Summary of incubation period estimates (unit: day) for cases infected by Omicron BA.1 variants and cases during Delta predominance period in South Korea

Type of SARS-CoV-2 Strain	sample size	estimate (95%CI)			
		mean	median	95-th percentile	SD
Omicron BA.1	$n = 22$	$4.1(2.5, 5.5)$	$3.8(2.4, 5.1)$	$7.9(4.3, 10.6)$	$2.0(1.0, 2.7)$
those during Delta dominance[$]	$n = 64$	$5.5(5.4, 8.0)$	$5.0(5.0, 6.4)$	$10.6(10.5, 15.5)$	$2.7(2.7, 3.9)$

[$] These cases were collected in June 2021 when the Delta variants were dominant at a prevalence of 68.3% in South Korea according to GISAID [42].

9.3 Change in disease burden associated with influenza and air pollutants during the COVID-19 pandemic in Hong Kong

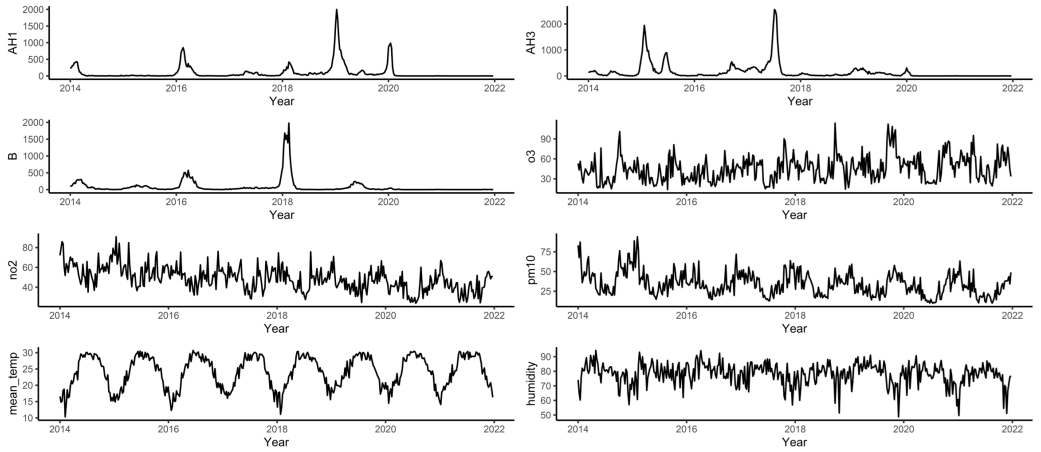

Figure 9.1: Time series plots of weekly proportions of influenza A (subtype H1N1, H3N2) and B, air pollutants in Hong Kong, 2014 to 2021

Table 9.9: Model goodness-of-fit and prediction accuracy.

Model	R^2	Adjusted R^2	MAPE	RMSE
XGBoost	0.939	0.937	2.056%	18.915
GAM	0.750	0.742	3.85%	37.662

Note: MAPE, mean absolute percentage error; RMSE, root-mean-square error.

Table 9.10: Overall excess mortality numbers and rates (per 100,000 population) of all-cause, cardiovascular diseases (CRD), pneumonia and influenza (P&I) by age group in 2020 and 2021, estimated from the GAM models, respectively.

	All-cause				CRD				P&I			
	Excess number	(95% CI)	Excess rate	(95% CI)	Excess number	(95% CI)	Excess rate	(95% CI)	Excess number	(95% CI)	Excess rate	(95% CI)
						2020						
All-age	6034	(4764, 7321)	80.97	(63.47,98.68)	3370	(2632, 4116)	45.28	(35.24,55.04)	679	(329, 1058)	9.16	(4.46,14.31)
0-19	41	(12, 72)	3.79	(1.1,6.55)	1	(-7, 11)	0.13	(-0.63,0.99)	-4	(-8, 2)	-0.34	(-0.78,0.23)
20-39	65	(13, 117)	3.30	(0.73,5.92)	35	(12, 58)	1.78	(0.62,2.96)	-2	(-11, 7)	-0.11	(-0.54,0.34)
40-64	665	(474, 857)	22.24	(16.04,28.73)	546	(437, 653)	18.20	(14.54,21.9)	43	(-3, 91)	1.45	(-0.02,3.04)
65-84	2444	(1840, 3020)	207.83	(156.39,256.21)	1469	(1109, 1822)	124.98	(93.68,156.12)	304	(147, 476)	25.89	(12.45,40.68)
85+	2199	(1539, 2897)	1005.32	(695.78,1331.95)	1107	(703, 1522)	507.95	(316,698.03)	318	(89, 541)	146.17	(40.71,251)
						2021						
All-age	9397	(8603, 10257)	126.24	(115.51,138.08)	5056	(4554, 5590)	67.86	(61.28,75.12)	987	(743, 1249)	13.30	(9.99,16.7)
0-19	22	(1, 42)	2.00	(0.11,3.92)	-1	(-7, 5)	-0.13	(-0.69,0.45)	-4	(-8, 1)	-0.41	(-0.78,0.05)
20-39	75	(28, 121)	3.91	(1.59,6.26)	29	(12, 47)	1.51	(0.63,2.43)	-1	(-10, 8)	-0.08	(-0.51,0.4)
40-64	1086	(913, 1242)	36.26	(30.64,41.5)	693	(593, 792)	23.08	(19.78,26.43)	16	(-26, 58)	0.52	(-0.89,1.94)
65-84	3648	(3226, 4058)	300.62	(266.24,333.11)	2240	(2004, 2475)	183.75	(164.11,202.98)	359	(228, 495)	29.54	(18.87,40.25)
85+	3579	(3107, 4084)	1581.25	(1374.19,1802.25)	1777	(1442, 2117)	785.68	(641.65,934.57)	572	(370, 773)	251.61	(164.12,338.54)

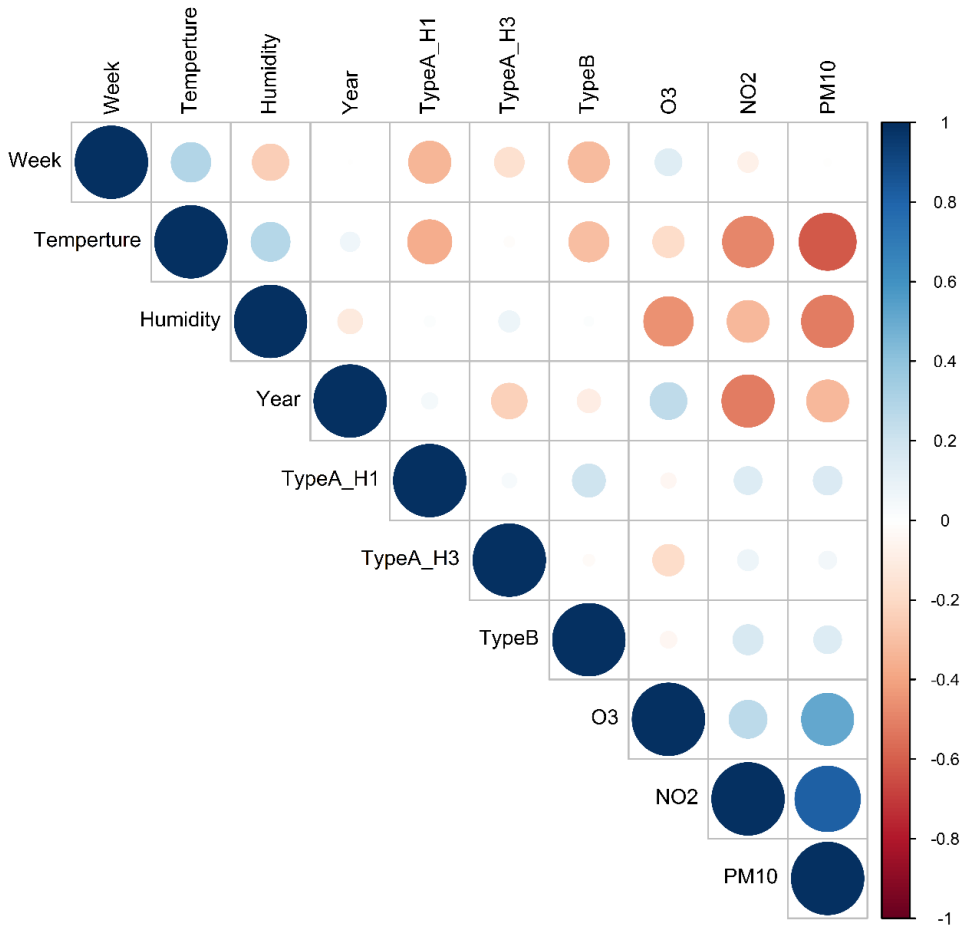

Figure 9.2: Correlation matrix of covariates.

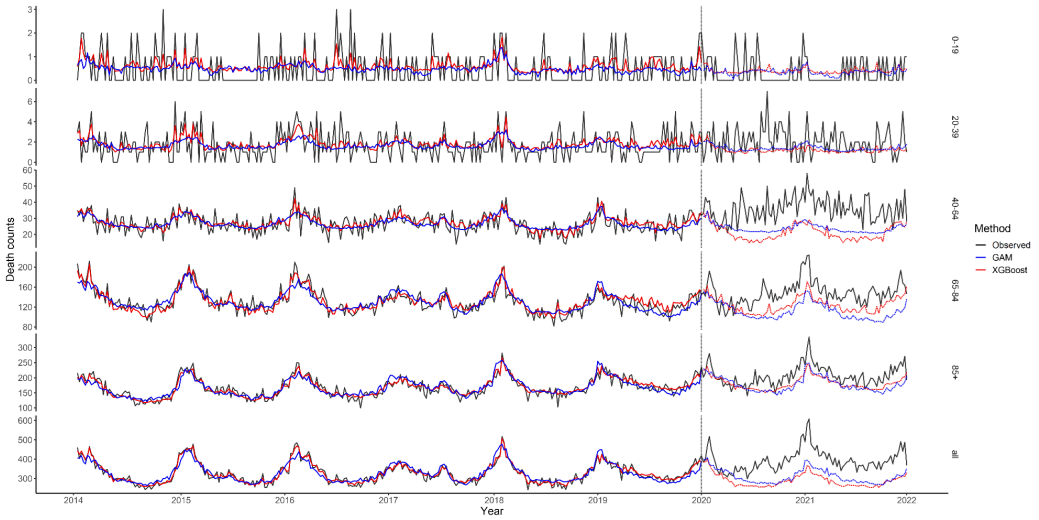

Figure 9.3: Observed and fitted weekly mortality data of cardiovascular and respiratory (CRD) by age groups. The XGBoost (blue line) and GAM models (red line) were developed from the training data from 2014 to 2019 and used to predict the data in 2020.

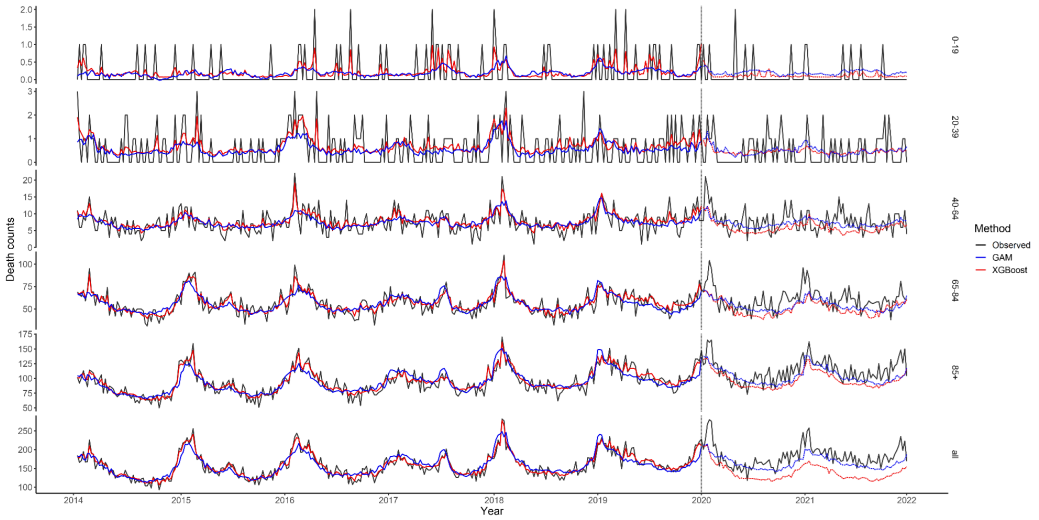

Figure 9.4: Observed and fitted weekly mortality data of pneumonia and influenza (P&I) by age groups. The XGBoost (blue line) and GAM models (red line) were developed from the training data from 2014 to 2019 and used to predict the data in 2020 and 2021.

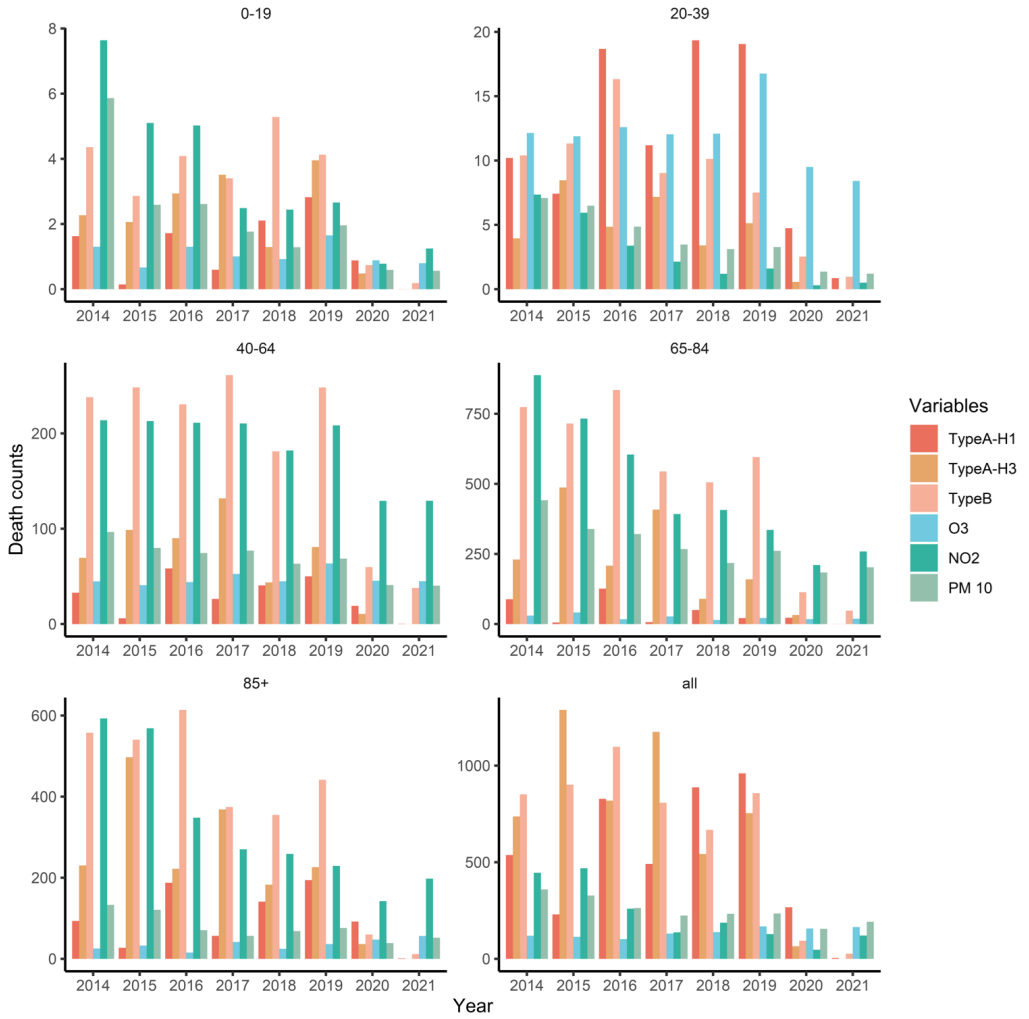

Figure 9.5: Annual CRD mortality attributable to influenza A (Flu A), influenza B (Flu B), O_3, NO_2, and PM_{10}, estimated from the XGBoost models.

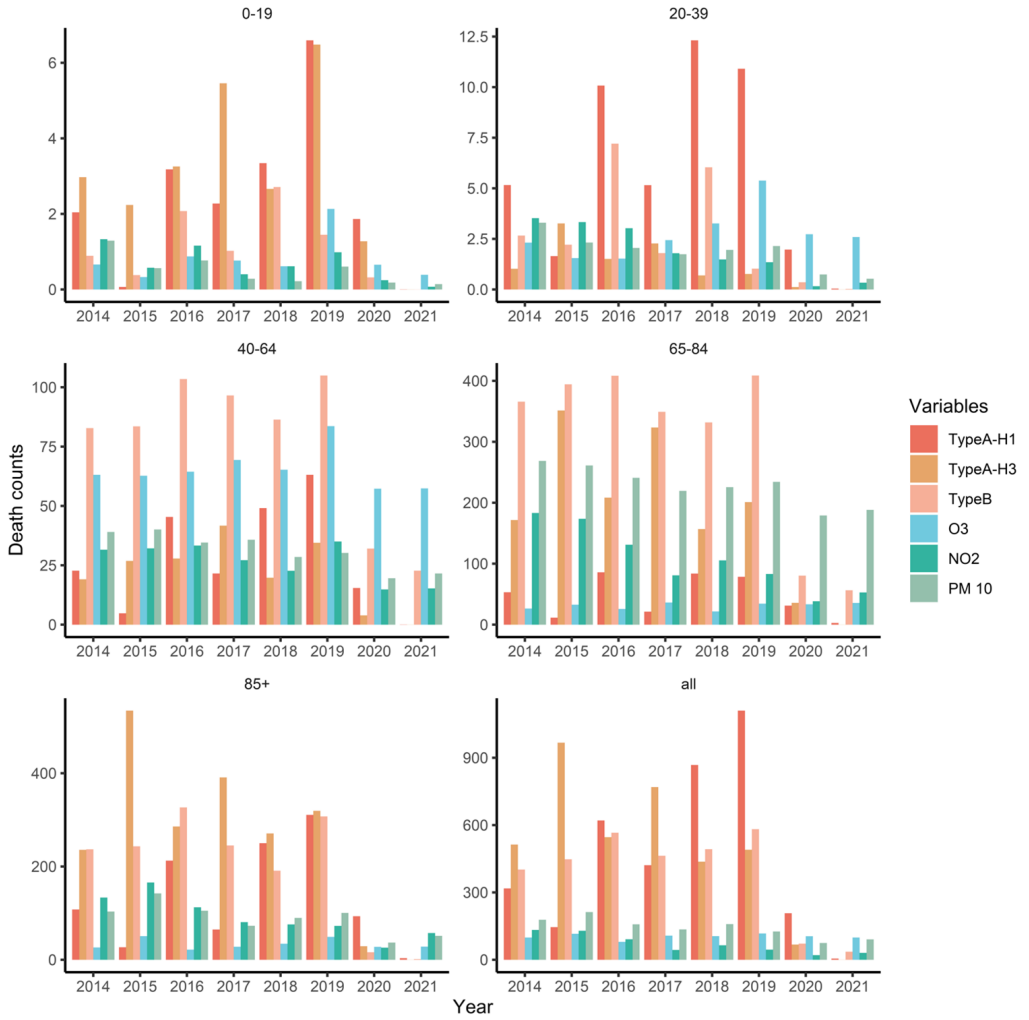

Figure 9.6: Annual P&I mortality attributed by major air pollution variables and influenza proxy estimated by the XGBoost models.

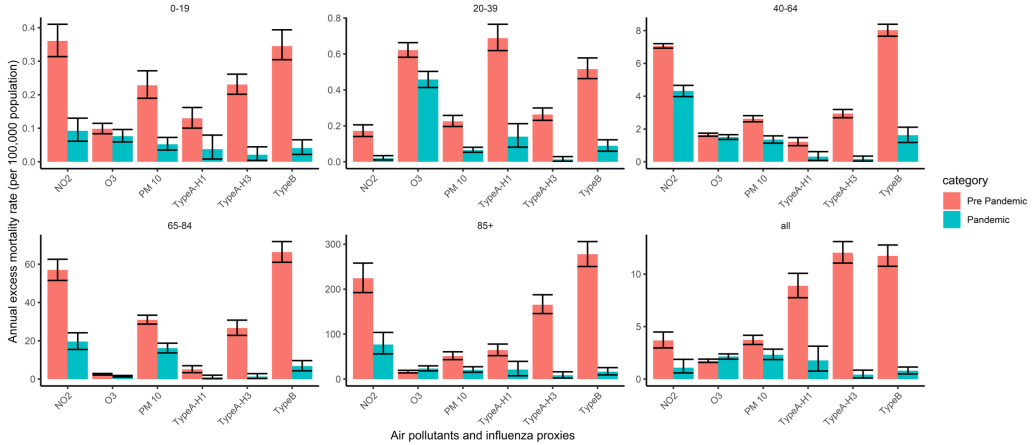

Figure 9.7: Annual CRD excess mortality rate (per 100,000 population) associated with air pollutants and influenza pre pandemic and during the COVID pandemic for different age groups with 95% confidence interval obtained from 10000 times bootstrap

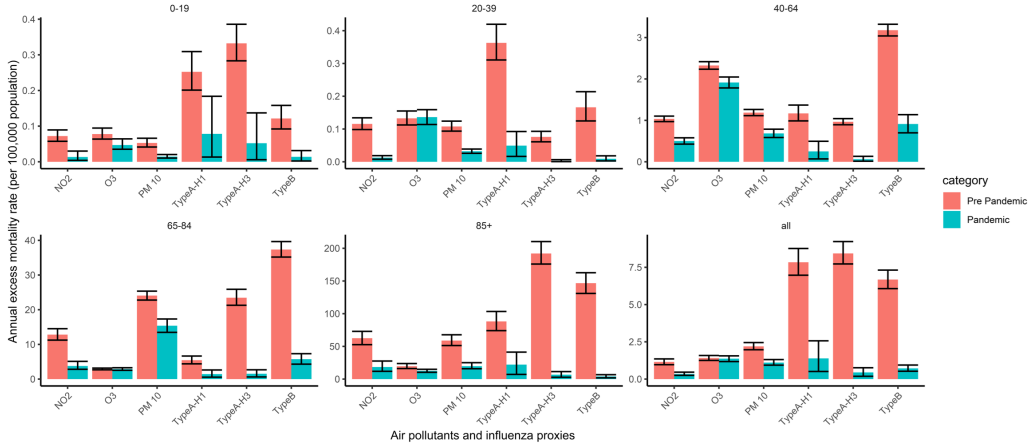

Figure 9.8: Annual P&I excess mortality rate (per 100,000 population) associated with air pollutants and influenza proxies pre pandemic and during the COVID pandemic for different age groups with 95% confidence interval obtained from 10000 times bootstrap.

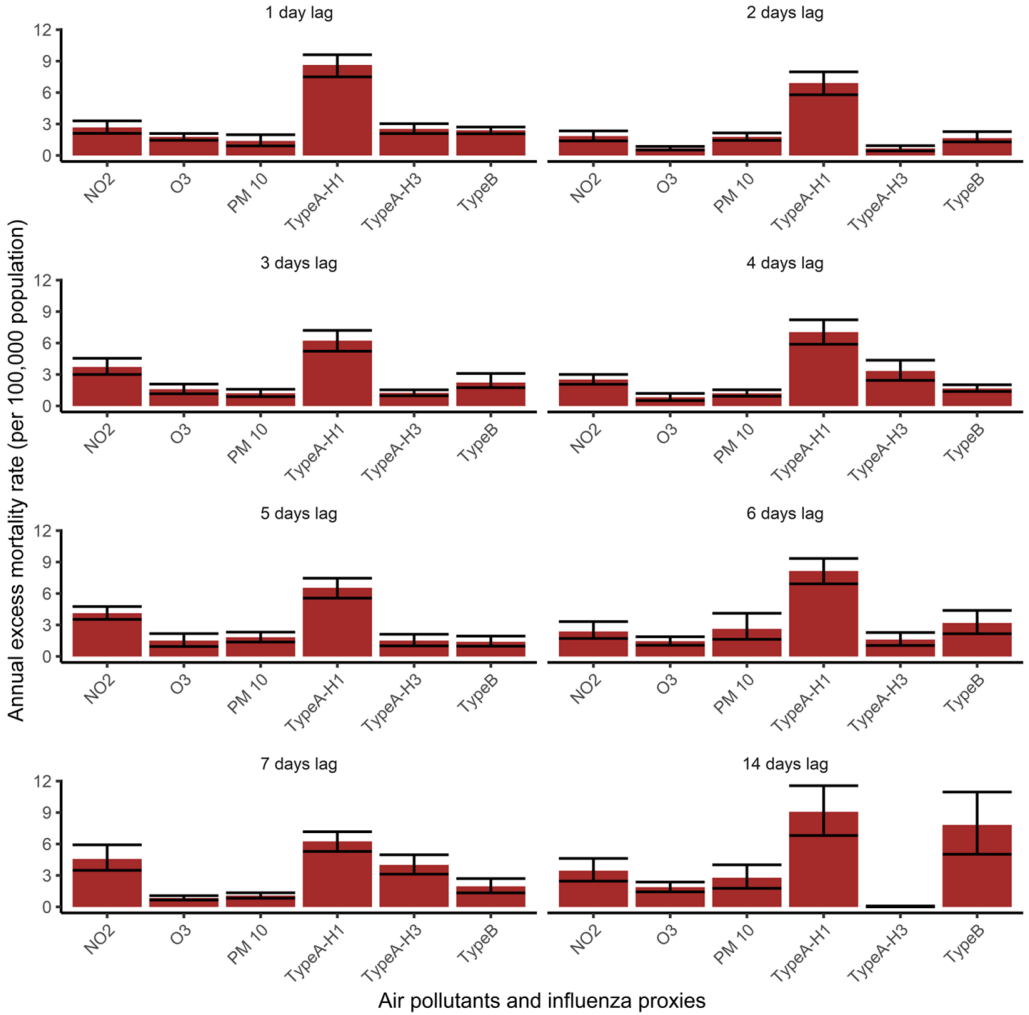

Figure 9.9: Lag effects (up to 14 days prior) for each influenza proxy and air pollutant. The estimates were from the XGBoost models.