Effect of long-term exposure to ambient fine particulate matter ($PM_{2.5}$) on the incidence of type 2 diabetes mellitus (T2DM): a cohort study in rural China

Cindy Yu

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Science degree in Epidemiology

School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa

ABSTRACT

Background: Long-term exposure to fine particulate matter ($PM_{2.5}$) has been identified as a potential risk factor for developing type 2 diabetes mellitus (T2DM). Given the rising prevalence of T2DM and unhealthy concentrations of $PM_{2.5}$ in China, our attention is brought to examining the association in this region of the world. Furthermore, rural China, although largely ignored, also finds itself suffering from increased risks of T2DM and high levels of $PM_{2.5}$.

Objective: The goal of this study is to characterize the relationship between long-term exposure to $PM_{2.5}$ and the risk of T2DM in rural China. We do so by confirming that greater long-term exposure to $PM_{2.5}$ is associated with a higher risk of T2DM incidence, assessing the potential multiplicative and additive interactions with important covariates, and identifying constituents of $PM_{2.5}$ that may be responsible for the effect $PM_{2.5}$ on the increased incidence of T2DM.

Conclusions: Greater long-term exposure to $PM_{2.5}$ is associated with increased risk of developing T2DM in rural Deqing County, Zhejiang, China. Smoking status modifies the relationship between $PM_{2.5}$ and T2DM incidence on a multiplicative scale. There is no synergism between smoking and $PM_{2.5}$ in association with T2DM incidence. There is no conclusive evidence on which constituents of $PM_{2.5}$ play greater roles in the adverse effects of $PM_{2.5}$ on T2DM incidence.

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Dr. Yue Chen, University of Ottawa, for your support and guidance throughout this chapter in my life. I feel lucky and blessed to have had such a thoughtful and understanding supervisor as you.

I would like to thank my thesis advisory committee member, Dr. Chaowei Fu, Fudan University, for your considerable feedback and engagement, and Dr. Haidong Kan, Fudan University, for providing your air pollution epidemiology expertise. Furthermore, I couldn't have done any of it without the help of your respective students, Yun Chen and Cong Liu, who, despite language barriers, made continuous efforts to answer my questions. To you all, thank you for welcoming me so warmly to Shanghai, thank you for helping me grow. I wish I had thanked you all more in person.

I couldn't mention warm welcomes to Shanghai without acknowledging my aunt and uncle, gūgu and gūfu! Thank you so much for opening your home and lives to me, I am so glad to have had the opportunity to become part of your family.

I would also like to thank my section team at Health Canada, for their patience and understanding whilst I work dually to contribute to their team and complete my thesis. I look forward to our time together from here on out.

And of course, I would like to thank my friends and family, for quietly supporting my studies in the background, and providing a sturdy foundation of support that I take for granted every day. I love you all and cherish our bonds so dearly. Finally, I would like to thank from deep within my heart, my lovely dog Penny, whose companionship brought me the greatest happiness I have ever known.

CONTENTS

List of Tables

List of Figures

List of Abbreviations and Acronyms

2hPG	2-hour plasma glucose
µm	Micrometers
µg/m^3	Micrograms per cubic meter
ACS CPS-II	American Cancer Society Cancer Prevention Study II Cohort
ADA	American Diabetes Association
AERONET	AErosol RObotic NETwork
AirGIS	Air pollution dispersion modeling system
AOD	Aerosol optical depth
APPCAP	Air Pollution Prevention and Control Action Plan
ATF4	Activating transcription factor 4
ATF6	Activating transcription factor 6
BAT	Brown adipose tissue
BC	Black carbon
BiP	Binding immunoglobulin protein
BME	Bayesian maximum entropy
BMI	Body mass index
BWHS	Black Women's Health Study
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CCHS	Canadian Community Health Survey
CDS	Chinese Diabetes Society
CHD	Coronary heart disease
China-PAR Project	Prediction for Atherosclerotic Cardiovascular Disease Risk in China Project
ChinaMUCA	China Multi-Center Collaborative Study of Cardiovascular Epidemiology
CIMIC	Community Intervention of Metabolic Syndrome in China & Chinese Family Health Study
CNBSS	Canadian National Breast Screening Study
CRP	C-reactive protein
CTM	Chemical transport models
DCCT	Diabetes Control and Complications Trial
eNOS	Nitric oxide synthase
EPA AQS	Environmental Protection Agency's Air Quality System
ER	Endoplasmic reticulum
EURAD-CTM	European Air Pollution Dispersion and Chemistry Transport Model
FPG	Fasting plasma glucose
GCLM	Glutamate-cysteine ligase modifier subunit
GEOS-CHEM	Goddard Earth Observing System Chemical Model
GIS	Geographic information system
GLUT4	Glucose transporter type 4
Gp96	Glucose regulatory peptide 94

HbA1c	Glycosylated hemoglobin; hemoglobin A1c
HNRS	Heinz Nixdorf Recall Study
HPFS	Health Professionals Follow-up Study
HR	Hazard ratio
Hsp70	Heat shock proteins 70
Hsp90	Heat shock proteins 90
ICAM-1	Intracellular adhesion molecule 1
ICD-9	International Classification of Diseases, Ninth Revision
ICD-10	International Classification of Diseases, Tenth Revision
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
IL-6	Interleukin 6
IL-10	Interleukin 10
IMPROVE	Interagency Monitoring of Protected Visual Environments
InterASIA	International Collaborative Study of Cardiovascular Disease in Asia
IQR	Interquartile range
IR	Incidence rate
IRR	Incidence rate ratio
IRS-1	Substrate 1
JKN	c-Jun N-terminal kinase
LUR	Land use regression
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NADPH	Nicotinamide adenine dinucleotide phosphate
NGSP	National Glycohemoglobin Standardization Program
NH_3	Ammonia
NH_4^+	Ammonium
NHS	Nurses' Health Study
NIH-AARP	National Institutes of Health and American Association of Retired Persons
NO	Nitric oxide
NO_3^-	Nitrate
NO_x	Nitrogen oxides; includes nitric oxide (NO) and nitrogen dioxide (NO_2)
NQO1	NAD(P)H quinone oxidoreductase 1
Nrf2	NF-E2–related factor 2
OC	Organic carbon
OGTT	Oral glucose tolerance test
ONPHEC	ONtario Population Health and Environment Cohort
OR	Odds ratio
p47	Cytosolic subunit of NADPH oxidase
PAI-1	Plasminogen activator inhibitor 1
PERK	Protein kinase-like endoplasmic reticulum kinase
PGC-1α	Peroxisome proliferator–activated receptor gamma coactivator 1 alpha

PI3-kinase/Akt	Phosphatidylinositol 3-kinase-Akt pathway
PM	Particulate matter
$PM_{2.5}$	Fine particulate matter; particles with aerodynamic diameters ≤2.5 μm
$PM_{1.0}$; UFP	Ultrafine particulate matter; particles with aerodynamic diameters ≤1.0 μm
PM_{10}	Coarse particulate matter; particles with aerodynamic diameters ≤10 μm
PPAR-γ2	Peroxisome proliferator-activated receptor gamma 2
Prdm16	PR domain containing 16
RoLS	Rome Longitudinal Study
RPG	Random plasma glucose
RR	Relative risk
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor
SES	Socioeconomic status
SO_2	Sulfur dioxide
SO_4^{2-}	Sulfate
SR-MA	Systematic review with meta-analysis
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TNF-α	Tumor necrosis factor alpha
TVE	Town and village enterprise
Ucp1	Uncoupling protein 1
UFP; $PM_{1.0}$	Ultrafine particulate matter; particles with aerodynamic diameters ≤1.0 μm
UPR	Unfolded protein response
VAT	Visceral adipose tissue
WAT	White adipose tissue
WHO	World Health Organization

CHAPTER 1. INTRODUCTION

1.1 Type 2 Diabetes Mellitus (T2DM)

There are two major types of diabetes mellitus, type 2 diabetes mellitus (T2DM) and type 1 diabetes mellitus (T1DM).[1] T2DM results when the body ineffectively responds to insulin whereas T1DM results when the body produces deficient amounts of insulin.[2] The majority of diabetes mellitus cases worldwide are T2DM.[3]

When not adequately managed, T2DM complications develop and can endanger the lives of those with T2DM. For instance, acutely, abnormally high blood glucose can trigger diabetic ketoacidosis and hyperosmolar coma, two life-threatening conditions. Abnormally low blood glucose can trigger seizures or loss of consciousness, commonly occurring after a meal is missed, more exercise than usually, or excessively high dosage of anti-diabetic medication. Chronically, T2DM can damage the heart, blood vessels, eyes, kidneys, and nerves, and increase the risks of heart disease and stroke. Reduced blood flow, in addition to nerve damage, can chronically lead to foot ulcers, limb amputation as well as diabetic retinopathy.[2]

Diagnosis of T2DM

Currently, there are four main diagnostic criteria for T2DM recommended by the World Health Organization (WHO) and the American Diabetes Association (ADA) using the following tests: fasting plasma glucose (FPG) test, 2-hour plasma glucose (2hPG) in an oral glucose tolerance test (OGTT), random plasma glucose (RPG) test, and glycosylated hemoglobin (HbA1c) test.[4–6] The diagnostic tests base their criteria on the measures of two main molecules: plasma glucose and HbA1c.[7]

In North America, T2DM is diagnosed if patients have: a FPG ≥126 mg/dL (7.0 mmol/L), for which fasting is defined as no caloric intake for at least 8 hours; 2hPG ≥200 mg/dL (11.1 mmol/L) during OGTT, performed using a glucose load of 75-g anhydrous glucose dissolved in water; RPG ≥200 mg/dL (11.1 mmol/L) in patients with classic symptoms of hyperglycemia or hyperglycemic crisis; or HbA1c ≥6.5% (48 mmol/mol), performed in a laboratory using a method certified by the National Glycohemoglobin Standardization Program (NGSP) and standardized to Diabetes Control and Complications Trial (DCCT) assay.[1,7]

All methods of diabetes diagnosis have advantages and disadvantages. FPG test is a single plasma glucose measurement completed after an overnight fast. It is advantageous for its simplicity, low cost, and low risk. However, it may be impractical, as it requires patients to fast, which has a possibility of being improperly done, or impeding promptness of testing for patients with afternoon appointments.[8] The OGTT is similarly disadvantageous, in that it requires extensive patient preparation, being inconvenient and time-consuming for patients.[9]

Largely due to this noted inconvenience in fasting for FPG, performing an OGTT, and the issue of day-to-day variability in glucose levels, HbA1c testing was sought as an alternative to measuring glucose. The major advantage of HbA1c is that it does away with the issues of day-to-day variability in glucose levels and does not require patients to make dietary preparations (i.e. fasting). However, a costly test, which is also not yet standardized worldwide, limits its use outside of high-income countries, such as Canada and the United States.[5,9] Furthermore, differences in cut-off are not clear for different races/ethnicities.[5] Evidently, the advantages and disadvantages of each diagnostic test must be weighed and considered carefully.

Given the ethnic differences in diabetes characteristics in East Asians compared to their Caucasian counterparts, for which ADA guidelines are designed for, China-specific guidelines were

necessary.[10] For instance, different genetic background may be responsible for different body-fat distribution in people of Asian descent, suggesting the need for different anthropometric reference values.[11]

Diagnostic tests used in Chinese surveys have changed over the last 20 years. In 1994, nationwide Chinese surveys tended to use OGTT alone, then FPG alone in 2000, OGTT with FPG in 2007, and OGTT with FPG and HbA1c in 2010.[12]

The Chinese Diabetes Society (CDS), founded in 1999, has been involved in the development and knowledge translation of diabetes guidelines and standards of care for the Chinese population; the most recent English version of the standards of care for T2DM in China was published in 2016.[10]

The 2016 guidelines for China recommended following the WHO's 1999 criteria for diagnosis and classification of diabetes.[13] Specifically, T2DM can be diagnosed using FPG, 2hPG OGTT, or typical symptoms of diabetes along with RPG.[10] None of the tests need retesting to confirm diabetes. Additionally, FPG, defined as measured blood glucose level at any time of day without fasting, can only be used for the diagnosis of diabetes but not for the diagnosis of prediabetes; venous plasma glucose from FPG or 2hPG OGTT alone may be used to diagnose prediabetes [i.e. impaired fasting glucose (IFG) or impaired glucose tolerance (IGT)]. The cut-off values to meet diagnostic criteria for diabetes in China follow the recommendations from the WHO and the ADA for FPG and 2hPG OGTT.

Although 2010 guidelines from the ADA allow for the use of HbA1c tests to diagnose T2DM,[6] also recommended by the WHO in 2011,[5] 2019 standards in China still do not recommend the use of HbA1c test for the diagnosis of diabetes in China: "Although the standardization of HbA1c test has been improved in China, it has not been sufficiently characterized to support routine adoption."[10] Despite this, the 2019 China standards encourage medical facilities and researchers to investigate the adequacy of HbA1c for diagnosis in their studies; several studies have suggested that a value of 6.3% may be the

optimal cut-off value for Chinese adults.[10,13] For now, only glucose-based tests are endorsed by the CDS.[14]

Known Risk Factors for T2DM

Over the years, epidemiologic research on T2DM has broadened our understanding of the major risk factors for the development of T2DM. Known risk factors for T2DM include family history of T2DM, obesity, poor diet, lack of physical activity, and smoking.[15,16] In China, where the population is constantly facing a rapid socioeconomic transition, such factors have also become major risk factors.[17] Family history of T2DM of having at least two first-degree relatives, one first-degree and one second-degree relative, or three second-degree relatives with T2DM drastically increases personal risk for T2DM.[18] However, fewer genetic factors have been identified in Chinese populations compared to Western counterparts, likely due to limited sample sizes of studies.[19] Increased urbanization has led many Chinese to be less physically active, as they move away from agrarian lifestyles. Consumption of sugar-sweetened drinks and higher calorie consumption have additionally increased rapidly in China.[20] The traditional Chinese diet, consisting of high consumption of rice, pork, and vegetables, is shifting towards a dietary pattern of high consumption of meats and edible oil, but low consumption of whole cereals and vegetables.[21] National nutrition surveys have shown that even in rural China, diets remain low in fruit and whole grains.[22] Altogether, such factors increase body fat; the WHO expert consultation in 2004 concluded that risk of T2DM starts at a body mass index (BMI) lower for Asians than their Western counterparts.[15] Moreover, in China, smoking prevalence is high and known to be a widespread for men; China is the greatest producer and consumer of cigarettes worldwide.[20] For smoking prevalence worldwide, China has one of the highest at 52.9% for men, and one of the lowest at 2.4% for women.[22]

Rapidly increased risk of T2DM Worldwide and in China

WHO estimates that 422 million adults aged 18 years or over were living with diabetes in 2014 globally.[2] Although such numbers cannot differentiate between cases of T2DM from T1DM, given that over 90% have T2DM,[23] and that our population of interest is adults, we make an assumption and refer to only T2DM. Worldwide, the number of cases of diabetes has increased drastically, from a prevalence of 4.7% in 1980 to 8.5% in 2014. The prevalence has risen even faster in low- and middle-income countries.[2] It is estimated that T2DM will have increased to 592 million adults worldwide by 2035, with the sharpest increases in low- and middle-income countries, such as China.[24]

In the 1980s, T2DM was not a major health issue in China.[25] In the past decade alone, the prevalence of diabetes has increased from 5.5% to 11.6%.[26] In a nationally representative epidemiological survey, Li et al. (2020) reported that the prevalence of T2DM in 2017 was 12.8% using the ADA diagnostic criteria or 11.2% using the WHO criteria.[27] Researchers associate China's increasing prevalence of T2DM with its rapid economic development, which has been accompanied by increased obesity, sedentary lifestyle, and poorer dietary choices.[25]

Researchers have realized that those major risk factors alone may not be responsible for the rise in T2DM prevalence. One important factor that potentially leads to the increased risk of T2DM is air pollution. Researchers have been long interested in assessing the associations of long-term exposure to air pollution with cardiovascular and respiratory diseases.[28] Only recently have researchers begun reporting findings on the association between air pollution and T2DM. In the 2017 Lancet Commission report on pollution and health, researchers outlined the important need to identify the causal links between air pollution and diseases including T2DM, particularly in low- and middle-income countries.[28]

1.2 Fine Particulate Matter ($PM_{2.5}$)

Growing epidemiologic research has widened our perspective on the range of risk factors for developing T2DM other than the major known risk factors, such as obesity, diet, and physical activity. One such risk factor is air pollution.[28–30] Particulate matter (PM) is most commonly classified by their size as coarse (PM_{10}), fine ($PM_{2.5}$), and ultrafine (UFP; $PM_{1.0}$). Fine particulate matter, hereon referred to as $PM_{2.5}$, is an important measure of ambient air pollution, and is defined as particles with an aerodynamic diameter of less than or equal to 2.5 micrometers (μm). $PM_{2.5}$ is measured in units of micrograms per cubic meter ($\mu g/m^3$).[31]

These particles of different size have different sources and formation mechanisms. For instance, some PM occurs naturally, such as dust storms and forest fires, whilst others are a product of human activities, such as changes in land use and fossil fuel combustion. In general, $PM_{2.5}$ is formed from non-combustibles in fuels, incompletely combusted fuels, and gas-to-particle conversion as sulfates, nitrates, ammonium, organic carbon, elemental carbon, heavy metals, and fine geological material. In comparison, UFP often take the form of carbon, sulfuric acid, and condensed metal vapours, and PM_{10}, as geological material, pollen, and sea salt. The major sources of $PM_{2.5}$ are from the combustion of coal, oil, gasoline, diesel fuel, and wood, the atmospheric transformation products of nitric oxides (NO_x), sulfur dioxide (SO_2), and organic compounds, as well as high-temperature processes, smelters and steel mills.[32]

The unique size of PM also leads to differences in how long they remain in the atmosphere and the distance they can travel. For example, the atmospheric half-life of $PM_{2.5}$ is days to weeks, whereas for PM_{10} it is minutes to hours.[32,33] Additionally, on average, $PM_{2.5}$ can travel great distances of hundreds to thousands of kilometers, but PM_{10} typically only travels less than ten kilometers.[32] For example, fine particles of black carbon, a form of $PM_{2.5}$, from burning of fossil fuels and biomass are capable of

travelling thousands of kilometers via the long-range transport process.[34] In addition, PM is classified as primary or secondary in terms of emission mechanisms. Primary PM is emitted directly from sources like burning, industry, and road dust. On the other hand, secondary PM is formed through chemical transformation of gaseous primary pollutants such as SO_2, NO_x, and ammonia (NH_3), among others.[32] Furthermore, these differences in size also lend themselves to differences in potential adverse health risks. Specifically, the smaller the particle size, the deeper into the breathing passages it can penetrate. Whereas PM_{10} can only pass into the throat and nose and enter the lung, $PM_{2.5}$ can further permeate into the alveolar region of the lung, presenting greater potential for adverse health effects.[24] WHO's Air Quality Guidelines recommend that the annual average $PM_{2.5}$ concentration should not exceed 10 $\mu g/m^3$, with interim targets starting at 35 $\mu g/m^3$.[35]

$PM_{2.5}$ Monitoring in China

The vast majority of epidemiologic studies on the associations between long-term exposure to ambient $PM_{2.5}$ and various health outcomes are based upon study populations in developed countries that have a long-history of quality fixed-site monitoring data and a vast, far-reaching network of monitors to provide $PM_{2.5}$ measurements. Advancements in technologies to map pollution, specifically satellite imaging, have opened the gates for examining associations between air pollution and a wider range of diseases than historically proposed. It is particularly important in regions of the world where there are a few or even no ground-level monitoring networks available.[28]

Given China's sudden and rapid industrialization and development, some areas of legislation put in place for air pollution prevention and control, which are long-standing practices in developed countries such as Canada and the US, had yet to catch fire in China. In particular, the Chinese government only began issuing policies regarding $PM_{2.5}$ in 2011, after citizens, whose concerns regarding

exposure to $PM_{2.5}$ were growing, publicly voiced their requests for better data availability. In addition to the pollutants already being monitored (i.e. SO_2, NO_x, and PM_{10}), the Chinese government amended the Ambient Air Quality Standards to include $PM_{2.5}$ as required in cities' pollution assessments.[36]

Compliance monitoring of $PM_{2.5}$ was first made mandatory for designated areas of Beijing, Shanghai, Guangzhou, and 31 capital cities, and over the next few years, designation of additional county-level cities to start monitoring $PM_{2.5}$ would be implemented.[37] By 2013, the Chinese government issued the Air Pollution Prevention and Control Action Plan (APPCAP) that led to the implementation of regulatory monitoring networks of $PM_{2.5}$ nationwide.[38]

Although this stringent air pollution control policy in response to public concerns is admirable, there is, nonetheless, a historical gap in available $PM_{2.5}$ monitoring data prior to 2013 that hinders investigations on the risk of adverse health outcomes due to long-term exposure to $PM_{2.5}$. Although satellite-derived $PM_{2.5}$ had been used to estimate surface-level $PM_{2.5}$ concentrations before, previously researchers relied on monitoring data provided by other countries, such as the US, to calibrate and validate satellite-derived $PM_{2.5}$ concentrations. Expectantly, there were concerns amongst experts regarding the applicability of US-specific monitoring data as training datasets for research in China. With the newly available nationwide monitoring $PM_{2.5}$ data in 2013, thereby providing China-specific calibration, researchers have even developed statistical models to estimate historical ambient $PM_{2.5}$ concentrations in China prior to 2013, allowing greater reaches to further explore the detrimental effects $PM_{2.5}$ on human health.[39]

Satellite-derived estimates of $PM_{2.5}$

Instruments onboard satellites passively retrieve aerosol optical depth (AOD) observations. Examples of instruments include: Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle

Imaging SpectroRadiometer (MISR), and Sea-Viewing Wide Field-of-View Sensor (SeaWiFS).[40–42] Chemical transport models (CTM) can simulate the ratio of AOD to $PM_{2.5}$, which can be applied to satellite-retrieved AOD to estimate surface-level $PM_{2.5}$. CTM can also be used to simulate the major chemical constituents of $PM_{2.5}$, such as inorganic aerosols [nitrate (NO_3^-), sulfate (SO_4^{2-}), ammonium (NH_4^+)], black carbon (BC), organic carbon (OC), and soil dust.[43] The Goddard Earth Observing System Chemical Model (GEOS-CHEM) is one CTM commonly used in conjunction with satellite imagery. Fixed-site monitors can aid in calibrating the satellite-retrieved AOD-$PM_{2.5}$ relationship by providing true measures of surface-level $PM_{2.5}$ and/or AOD and be used in a training dataset. The AErosol RObotic NETwork (AERONET), which measures surface-level photometry AOD, is a monitor that is commonly used for calibration. Surface-level *in situ* observations of $PM_{2.5}$ can also be used to further validate how well satellite-derived $PM_{2.5}$ approximates surface-level $PM_{2.5}$. Examples of networks of in situ observations commonly used in validation studies of satellite-derived $PM_{2.5}$ include the Interagency Monitoring of Protected Visual Environments (IMPROVE) and the Environmental Protection Agency's Air Quality System (EPA AQS). Other satellite instruments have been used to improve correlation between satellite-retrieved column AOD and surface-level AOD by providing active retrievals for vertical profile adjustment, such as the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP).[40–42]

Advancements in satellite imaging data to be used in mapping air pollution exposure has allowed us to examine the association between air pollution and disease in regions with limited monitoring data.[28,31,44]

1.3 Rural China

Rural China presents a unique region often misunderstood and neglected by research, and yet would greatly benefit the attention, as there exist important behavioural and demographical differences

between urban and rural China that likely impact the generalizability of most findings to this population. Most studies are performed in developed countries that have many years of monitoring data, allowing for long-term effects to be readily researched. Furthermore, of the limited research done in China, most take place in urban megacities. Therefore, the etiology of T2DM in rural populations greatly warrants further research and attention.

T2DM in Rural China

Many cases of diabetes remain undiagnosed in China.[45] Additionally, of those who are diagnosed, management is lacking, particularly in rural areas, leading to greater risk of premature death.[45]

Bragg et al. (2017) examined the association between diabetes and cause-specific excess mortality in rural and urban areas in China and found that the prevalence of diabetes was higher in urban than rural areas, which was consistent with previous studies; however, individuals with diabetes in rural areas had more than twice the risk of all-cause mortality compared to their urban counterparts.[45] Furthermore, the risks of deaths from stroke, diabetic ketoacidosis or coma, and chronic kidney disease, were much greater than what is indicated in reports from high-income countries, and were particularly greater in rural areas of China.[45] These main causes of death associated with diabetes differ from Western populations, where death by coronary heart disease (CHD) is more common than for stroke.[45] Importantly, these poor prognostic outcomes associated with diabetes are likely a result of poor clinical management that is common in rural China.[25]

Healthcare accessibility is lacking in rural areas, and often are centrally located in the most well-developed region in each county.[46] It has even been suggested that rural village doctors often lack necessary knowledge and skills due to never receiving formal training in diabetes care. Furthermore, the

health system itself does not incentivize knowledge acquisition and prevention of diabetes. For example, in Anhui province, prescription of medication to treat diabetes can only be done by doctors at township or higher levels, not village doctors.[47]

Moreover, a study by Liu et al. (2016) that combined epidemiological research and meta-analysis, found that awareness, treatment, and control of T2DM in rural Chinese population showed much lower proportions compared to those in developed countries.[48]

$PM_{2.5}$ in Rural China

Exposure to $PM_{2.5}$ is not solely an urban problem.[49] For instance, in the process of transitioning into a market economy, enterprises and local governments focused their resources on performance rather than pollution and associated health issues. Town and village enterprises (TVEs) have been found to discharge higher levels of pollutants in these poor rural areas, where enforcement of regulations are weak.[50] This gap in environmental inequality between the rural and urban areas is evident in many instances. Many pollution enterprises in China, such as coal-fired power plants, industries and factories, are moved out of urban areas and into rural ones. For example, in order to improve the air quality in Beijing for the Olympic Games 2008, all pollution enterprises were moved out of the Fifth Ring Road and into rural regions.[50] Furthermore, it is well known that urban areas such as Beijing have extensive concentrations of air pollution, but significant levels extend throughout the country; this is largely expected because PM can remain airborne for days to weeks and can travel for many thousands of kilometers.[32,51,52]

China has implemented many monitoring systems as part of the Chinese National Real-time Air Quality Network, but the air quality network is skewed towards urban regions, which often have several

sites per city, and one or none in rural regions.[51] Therefore, use of satellite-derived air pollution data is essential in adequately assessing the health effects in rural regions of China.

1.4 Biological Mechanisms

The importance of evidence for biologic plausibility is essential, as epidemiology on its own, cannot establish causality.[53] Although it is common knowledge that a genetic predisposition, poor dietary and sedentary behaviour choices, and obesity are major players in the development of T2DM, researchers suggest that those factors alone are not the only drivers of the rapid increase in T2DM prevalence.[54] Recent epidemiological and biomechanism research suggests that long-term exposure to $PM_{2.5}$ may have detrimental effects on T2DM. $PM_{2.5}$ can deposit deep within the lungs, at the alveoli-level, and interact with and modify cellular- and molecular-level structures.[55] From *in vivo* and *in vitro* studies, researchers have hypothesized four major mechanisms by which $PM_{2.5}$ may lead to T2DM: endothelial dysfunction, oxidative stress and systemic inflammation, hepatic insulin resistance and stress, and mitochondrial and adipose tissue dysfunction.[3,54,56–60]

Vasculature

The endothelium is the innermost lining of the blood vessel, meaning that it has direct contact with the bloodstream. Endothelium dysfunction is defined by impaired ability to maintain vascular homeostasis.[61] Endothelial dysfunction is known to precede changes in insulin resistance and lead to reduction in peripheral glucose uptake.[54] Proinflammatory mediators attenuate the signalling of insulin by impairing the tyrosine kinase activity of the insulin receptor.[61] $PM_{2.5}$ exposure may lead to elevated levels of proinflammatory mediators [e.g. tumor necrosis factor alpha (TNF-α), interleukin 6 (IL-6), and resistin].[3] $PM_{2.5}$ exposure may also lead to elevated levels of prothrombotic adipokines [e.g.

plasminogen activator inhibitor 1 (PAI-1)], also known to mediate endothelial dysfunction. Adhesion molecules [e.g. intracellular adhesion molecule 1 (ICAM-1) and E-selectin] are important in promoting leukocyte adherence in the postcapillary venular endothelium; $PM_{2.5}$ exposure may lead to elevated levels of circulating adhesion molecules.[3]

In addition to acting as a mediator of endothelial dysfunction, $PM_{2.5}$ may also lead to abnormal insulin in the vasculature. Insulin normally activates the phosphatidylinositol 3-kinase-Akt pathway (PI3-kinase/Akt) in endothelial cells to activate endothelial nitric oxide synthase (eNOS), leading to the release of nitric oxide (NO), which in turn, improves the flow of blood and improves uptake of glucose.[61] However, $PM_{2.5}$ exposure has shown to impair the signalling of PI3-kinase/Akt eNOS in the aorta and to decrease the tyrosine phosphorylation of insulin receptor substrate 1 (IRS-1) in the liver.[3]

Immune System

It is widely believed that immune activation or inflammation is an important mechanism for obesity increasing the risks of insulin resistance and T2DM.[56] $PM_{2.5}$ exposure may be associated with elevated systemic proinflammatory biomarkers [e.g. C-reactive protein (CRP), IL-6, fibrinogen, white blood cells, circulating soluble adhesion molecules].[62] $PM_{2.5}$ exposure may lead to increased number of macrophages in adipose tissue, with a shift from the normal phenotype to the proinflammatory phenotype, characterized by increased expression of TNF-α and IL-6 genes and decreased expression of interleukin 10 (IL-10) gene].[3,62] A similar switch (i.e. an increase in innate immune cells) in the visceral adipose tissue (VAT) is believed to be a prominent feature of T2DM. $PM_{2.5}$ exposure may additionally direct the innate immune response in VAT by increasing the phosphorylation of cytosolic subunit (p47) of NADPH (nicotinamide adenine dinucleotide phosphate) oxidase leading to oxidative stress and subsequent insulin resistance.[3]

Liver

Another hallmark feature of T2DM is defective insulin signalling in the liver. Insulin normally stimulates the liver to store excess glucose as glycogen. Phosphorylation of the receptor activates the catalytic activity of the receptor. Once activated, the receptor then phosphorylates intracellular proteins, generating a response. When IRS-1, an insulin receptor in the plasma membrane, receives insulin and activates via phosphorylation, this leads to activated PI3-kinase/Akt signalling and glucose transporter type 4 (GLUT4) translocation. GLUT4 is the major transporter for glucose uptake in the plasma membrane, which are present in the cytoplasmic vesicles when there is no insulin and are ultimately unusable in their location. When insulin binds IRS-1, GLUT4 is translocated from the vesicles to the plasma membrane, where it can enact its duty of glucose uptake.[63] $PM_{2.5}$ exposure leads to decreased phosphorylation of Akt in the liver and skeletal cells, deposition of hepatic lipids, and decreased gluconeogenesis. This decrease in phosphorylation leads to defective PI3-kinase/Akt signalling, thereby suppressing the translocation of GLUT4 and resisting the effects of insulin's presence.[3]

In addition, $PM_{2.5}$ exposure may lead to endoplasmic reticulum (ER) stress. When the homeostasis of the ER is impaired, it has a mechanism called unfolded protein response (UPR), used to alleviate protein misfolding in response to adverse stressors.[60] Chronic hyperglycemia and hyperlipidemia, important factors leading to T2DM, are known to disrupt homeostasis of ER and induce the activation of the UPR mechanism.[64] $PM_{2.5}$ may also lead to elevated levels of UPR-associated proteins [e.g. activating transcription factor 4 (ATF4), heat shock proteins 70 (Hsp70) and 90 (Hsp90), and binding immunoglobulin protein (BiP)].[3]

One of the main indicators for ER stress is the activation of the activating transcription factor 6 (ATF6) pathway; $PM_{2.5}$ exposure may lead to elevated levels of glucose regulatory peptide 94 (Gp96) and

BiP in the lungs and liver, an indication of activation of the ATF6 pathway. Another main indicator for ER stress is the double-stranded RNA-activated protein kinase-like ER kinase (PERK) in the liver. ER stress, along with activation of the UPR, interacts with multiple inflammatory and stress signalling pathways, which additionally leads to changes in glucose metabolism. $PM_{2.5}$ exposure may lead to activation of inflammatory pathway through c-Jun N-terminal kinase (JNK), downregulation of signalling mediated by IRS-1, and downregulation of expression of peroxisome proliferator-activated receptor γ2 (PPAR-γ2) in the liver.[3]

Adipose Tissue

Obesity is a well-known risk factor for developing insulin resistance and T2DM. There are two types of adipose tissue, white adipose tissue (WAT) and brown adipose tissue (BAT). WAT is the main resource for energy and is responsible for secreting hormones and cytokines that regulate metabolism and insulin resistance. BAT is responsible for regulating non-shivering thermogenesis to expend energy. Obesity development is not only a function of food intake and energy expenditure, but additionally a balanced ratio of WAT to BAT.[60,65] There has been evidence of a relationship between obesity, BAT, and insulin resistance. Those who are obese have defective metabolism of fatty acids via β-oxidation in the mitochondria, which leads to an accumulation of intracellular metabolites in the skeletal muscle and liver (e.g. fatty-acyl CoA, diacylglycerol, and ceramide). Mitochondrial damage is a known contributor to insulin resistance and T2DM. Smoking cigarettes and having hypercholesterolemia additionally lead to similar mitochondrial damage and subsequent accumulation of intracellular metabolites. Long-term $PM_{2.5}$ exposure may lead to decreased amount of interscapular BAT and decreased mitochondrial size, along with increased oxidative and nitrosative stress in BAT and induction of phase II antioxidant genes [e.g. NF-E2–related factor 2 (Nrf2), NAD(P)H quinone oxidoreductase 1 (NQO1), and glutamate-cysteine

ligase modifier subunit (GCLM)]. Additionally, $PM_{2.5}$ exposure may lead to decreased expression of uncoupling protein 1 (Ucp1) and peroxisome proliferator–activated receptor γ coactivator 1α (PGC-1α) in BAT, but decreased expression of PR domain containing 16 (Prdm16), PGC-1α, and PPAR-γ2 in WAT. The downregulation of these pathways specifically in BAT compared to WAT is suggestive that long-term $PM_{2.5}$ exposure may lead to altered insulin sensitivity in adipose tissue. Even in different models of insulin resistance, downregulation of these other insulin resistance pathways in BAT and WAT similarly have suggested that $PM_{2.5}$ modulates insulin sensitivity in adipose tissue.[3]

1.5 Research Objectives

The main purpose of this thesis was to examine the relationship between long-term exposure to $PM_{2.5}$ and the incidence of T2DM based on data from a rural cohort in China. Specifically, the research objectives were:

1. To confirm whether greater long-term exposure to $PM_{2.5}$ is associated with increased risk of developing T2DM;
2. To assess whether additional individual- or neighbourhood-level covariates had noticeable influence on $PM_{2.5}$ exposure in association with T2DM incidence synergistically or antagonistically; and
3. To identify which $PM_{2.5}$ constituent may play a major role in the deleterious effects of $PM_{2.5}$ on T2DM.

CHAPTER 2. LITERATURE REVIEW

Interest in examining the association between long-term exposure to $PM_{2.5}$ and T2DM has been on the rise. During the period of 2014–2017, there were eight systematic reviews with meta-analysis (SR-MA) published in English.[26,66–72] Within the last two years, three additional SR-MAs on the association between air pollution and T2DM have been published,[73–75] suggesting widespread interest on the topic. Details on the study population characteristics and results were extracted from the cohort studies of long-term $PM_{2.5}$ exposure and incidence of T2DM, and a summary table can be found in **Appendix 1** (**Table A1.1**). Details about different cohort characteristics, exposure metrics, and study follow-up can be found in the form of a heat map in **Appendix 1** (**Table A1.2**). Most authors chose to present measures of effect [e.g. hazard ratio (HR), odds ratio (OR)] per 10 $\mu g/m^3$ increase in $PM_{2.5}$; however, a still substantial number of authors chose to present their results per interquartile range (IQR) increase in $PM_{2.5}$. To facilitate comparability, results presented per IQR were converted to per 10 $\mu g/m^3$.

2.1 North America

There were 13 cohort studies done in North America, of which six were from Canada.[76–81] Two studies examined the association at the national scale,[77,81] and the remaining four examined the association at a smaller regional scale.[76,78–80]

Paul et al. (2020) examined the T2DM incidence in a population-based cohort, the ONtario Population Health and Environment Cohort (ONPHEC), in Ontario, Canada. Eligible adults aged 35 to 85 years at baseline were followed up from 2001 to 2015. Ground-level ambient exposure was approximated using satellite-derived estimates at a 1×1 km spatial resolution. Personal exposure was assigned to each individual as a 3-year moving average and linked using annual residential postal codes,

thereby accounting for residential mobility. The overall mean concentration of $PM_{2.5}$ was 9.6 μg/m^3 and the IQR was 3.5 μg/m^3. Incidence was identified from the Ontario Diabetes Database and defined using International Classification of Diseases (ICD)-9 or ICD-10 codes. Of the 4,774,984 individuals free of disease at baseline, 790,461 incident cases were identified during follow-up. After adjustment for potential confounding by age, sex, socioeconomic status (SES), region, and healthcare access, Paul et al. (2020) found that a 10 μg/m^3 increase in $PM_{2.5}$ exposure was associated with a 3% increase in T2DM incidence (HR = 1.03, 95% CI: 1.02, 1.04).[76]

Requia et al. (2017) investigated adults aged 44 years or older at baseline from all over Canada, from 2007 to 2014 using data from the Canadian Community Health Survey (CCHS). Ground-level $PM_{2.5}$ concentration was approximated using satellite-based estimates at a 1×1 km spatial resolution. Exposure was assigned as a 5-year time-weighted average and linked to individuals by health region, rather than individual-level linkages such as residential postal codes. They did not report overall regional concentrations of $PM_{2.5}$. Of the total sample size of 25,704,288 Canadians, they identified 5,570,326 cases during follow-up. The incidence rate ratio (IRR) was 1.05 per 10 μg/m^3 increase in $PM_{2.5}$ exposure (95% CI: 1.02, 1.13) after adjusting for age, cohort, and gender.[77]

Clark et al. (2017) conducted a cohort study in British Columbia, Canada. In smaller regional area, the Metropolitan of Vancouver, they developed a population-based cohort of adults aged 45 to 85 years at baseline. The study period was from 1994 to 2002, with four years of follow-up on average. Although additionally done at a 10×10 km spatial resolution, Clark et al. (2017) utilized high-spatial-resolution land use regression (LUR) to estimate exposure. Five-year time-weighted average $PM_{2.5}$ was assigned to each participant linked using annual residential postal codes. Incidence was defined using ICD-9 or ICD-10 codes and identified with the province's Medical Services Plan, which includes central registry, physician visit, and hospital data. The overall regional mean $PM_{2.5}$ was 4.1 μg/m^3, ranging from 0.0 to 10.2 μg/m^3, with an IQR of 1.6 μg/m^3. Of a total 380,738 participants without T2DM at baseline,

12,941 cases were identified during follow-up. After adjusting for age, gender, and area-level household income, a 10 μg/m^3 increase in $PM_{2.5}$ exposure was significantly associated with a 20% increase in T2DM incidence (OR = 1.20, 95% CI: 1.07, 1.36).[78]

To et al. (2015) and Chen et al. (2013) reported similar findings, in two different cohort studies both done in Ontario, Canada.[79,80] To et al. (2015) studied women aged 40 to 59 years at baseline who participated in the Canadian National Breast Screening Study (CNBSS) in 1992 and followed them up from 1996 to 2013. Like in the study done by Paul et al. (2020), exposure was approximated using satellite-derived estimates but used a less precise spatial resolution (10 × 10 km). Personal exposure was assigned using residential postal codes at baseline only; residential mobility was therefore unaccounted for. Cases of T2DM were defined using ICD-9 or ICD-10 codes and identified from administrative data, including physician outpatient claims, emergency visits, and hospital discharges. After adjustment for age, education, occupation, marital status, smoking, BMI, and other contextual variables, each 10 μg/m^3 increase in exposure was significantly associated with 28% increased risk of developing T2DM (IRR = 1.28, 95% CI: 1.13, 1.45).[79]

Chen et al. (2013) conducted another population-based cohort study of adult residents of Ontario, Canada who were 35 years old or older at baseline. Similar to To et al. (2015), they approximated ground-level exposure using satellite-derived estimates at a 10 × 10 km spatial resolution. Personal exposure was assigned as a 6-year average and linked to individual residential postal codes at baseline. The $PM_{2.5}$ concentration over the study region was 10.6 μg/m^3 on average and ranging from 2.6 to 19.1 μg/m^3. Incident cases were identified using the same sources as used in To et al. (2015)'s study, and it was indicated that the registry and algorithm were validated. Over the study period from 1996 to 2010, the mean follow-up was 8 years (±3.2 years). Of the total 62,012 participants, 6,310 cases were identified during follow-up; the incidence rate (IR) was 13.0 per 1,000 person-years. The adjusted HR was 1.11 (95% CI: 1.02, 1.21) per 10 μg/m^3 increase in exposure.[80]

Brook et al. (2013) examined the relationship in a Canadian-wide population-based cohort, called the Canadian Census Mortality Follow-up Study. Eligible participants were slightly younger (25 years old or older at baseline) and were followed up from 1991 to 2001. They used satellite-based estimates at a wide spatial resolution (10 $\times$ 10 km). Exposure was assigned as a 6-year average and linked to individual residential postal codes at baseline. Uniquely, mortality due to T2DM was the measurable outcome of interest and was defined using ICD-9 or ICD-10 codes identified in the Canadian Mortality Database. Over the approximately 10 total years of follow-up, 5,200 of the 2,145,400 participants were identified as having a T2DM-related death. The regional mean $PM_{2.5}$ concentration was 8.7 (±3.9) $\mu g/m^3$ and the IQR was 6.2 $\mu g/m^3$. After adjusting for potential confounders, each 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure was associated with a 49% increase in T2DM mortality (HR = 1.49, 95% CI: 1.37, 1.62).[81]

Expectantly, as Canada and the United States are both situated on the North American continent and additionally present comparable anthropomorphic polluting behaviours, evident in that both consistently maintain $PM_{2.5}$ values averaged <10 $\mu g/m^3$,[31,33] the current findings from cohort studies on long-term exposure to $PM_{2.5}$ and development of T2DM are similar. Furthermore, human time-activity patterns and behaviours are likely to be comparable between the two countries. There were seven cohort studies done in the United States.[82–88] Five papers examined the association at the national scale[82,83,85,86,88] and two studies,[84,87] done by the same authors, examined the association in Los Angeles.

Lim et al. (2018) developed a nationwide cohort study of participants who were aged 50 to 71 years at baseline of the National Institutes of Health and American Association of Retired Persons (NIH-AARP) Diet and Health Study. Over a follow-up period from 1995 to 2011, they assigned EPA AQS monitoring $PM_{2.5}$ data to each participant as a time-fixed 8-year average, including assignment as a time-varying exposure with 1-year lag as a sensitivity analysis. The overall regional mean concentration of

$PM_{2.5}$ was 11.0 (±2.7) μg/m^3. They assessed the relationship between long-term exposure and mortality of T2DM, from Social Security Administration Death Master Files, and collected data on additional covariates from mailed questionnaires. After adjusted for confounding variables, the authors found that the HR per 10 μg/m^3 increase in $PM_{2.5}$ was 1.19 (95% CI: 1.03, 1.39). In their sensitivity analysis using time-varying exposure, the HR was comparable at 1.18 (95% CI: 1.03, 1.36). The authors also found that higher BMI and lower fruit consumption positively modified the effect of $PM_{2.5}$ exposure on T2DM mortality.[82]

Bowe et al. (2018) developed a nationwide cohort study of US veterans, of whom 94% were men. The study lasted from 2003 to 2012, with a median follow-up of 8.5 years. They used two methods of exposure assessment: first they used EPA AQS monitors within 30, 10, and 5 miles from annual residential postal codes, and then alternatively they used satellite-based estimates. Moreover, they chose to assign exposure in their statistical model as either baseline concentration (time-invariant variable) or as annual concentration (time-varying variable). Incidence of T2DM was identified from the US Department of Veterans Affairs' databases, which is composed approximately 94% of men, and includes extensive data on routine blood panels. T2DM was defined using ICD-9 codes, diabetes medication prescription, or HbA1c >6.4% (>46.4 mmol/mol). The overall regional mean concentration of $PM_{2.5}$ was 1.8 μg/m^3, ranging from 5.0 to 22.2 μg/m^3, and the IQR was 7 μg/m^3. There were a total of 1,729,108 Americans with residential postal codes within EPA monitoring regions and 1,670,031 within satellite regions. A total of 397,966 cases were identified during follow-up, resulting in an incidence rate (IR) of 34.1 per 1,000 person-years. After adjustment for potential confounders, the adjusted HR per 10 μg/m^3 increment was 1.15 (95% CI: 1.08, 1.22) when assigned as baseline exposure and 1.18 (95% CI: 1.10, 1.25) when assigned as time-varying exposure. The HR using satellite-derived estimates was comparable at 1.13 (95% CI: 1.11, 1.15).[83]

The study by Park et al. (2015) in the US consisted of adults aged 45 to 84 years at baseline from multiple ethnicities. Over a 9-year follow-up period, they identified 622 T2DM cases out of the total 5,135 study participants. They used a hierarchical spatiotemporal model with collections from the EPA AQS and supplemented with LUR data to quantify $PM_{2.5}$ concentration, and assigned values at residential address as annual average $PM_{2.5}$ exposure at baseline. The overall mean $PM_{2.5}$ was 16.7 $\mu g/m^3$ with an IQR of 2.43 $\mu g/m^3$. Incident T2DM was defined from clinical examinations using the definition of FPG levels greater than or equal to 126 mg/dl, or a doctor diagnosis or taking diabetes medication. A questionnaire was administered to obtain data on additional covariates. The IR was 15.9 per 1,000 person years. After adjustment for potential confounders, the HR per 10 $\mu g/m^3$ increase in $PM_{2.5}$ was 1.08 (95% CI: 0.79, 1.49). The authors found no significant effect modification by gender, ethnicity, education, or obesity.[85]

Pope et al. (2015) developed a cohort study using participants of the American Cancer Society Cancer Prevention Study II Cohort (ACS CPS-II), who were aged 30 or older at baseline. Of the total 669, 046 participants who were followed up from 1982 to 2004, 4,890 died of T2DM. They used a hybrid LUR and Bayesian maximum entropy (BME) model to estimate $PM_{2.5}$ concentrations, which were assigned to participants as a time-fixed 5-year average. The overall mean $PM_{2.5}$ was 12.6 (±2.9) $\mu g/m^3$, ranging from 1 to 28 $\mu g/m^3$. T2DM mortality data came from a national database that includes details of biennial vital status on death certificates. Additional covariates were collected using a baseline questionnaire. The adjusted HR per 10 $\mu g/m^3$ increase in $PM_{2.5}$ was 1.13 (95% CI: 1.02, 1.26).[86]

Puett et al. (2011) conducted analyses on two separate cohorts, one made up of women aged 30 to 55 years at baseline from the Nurses' Health Study (NHS), and the other made up of men aged 40 to 75 years at baseline from the Health Professionals Follow-up Study (HPFS). Both cohorts used the same geographic information system (GIS)-based spatiotemporal models to quantified $PM_{2.5}$ concentration and assigned $PM_{2.5}$ as time-varying exposure with 1-year lag to each participant. Of note,

while both cohorts linked exposure by residential postal code at baseline, the HPFS cohort also sometimes linked exposure by work postal codes. Incidence of T2DM was identified via self-reported biennial mailed questionnaires, which were validated with medical records. The overall mean $PM_{2.5}$ was comparable, being 17.5 (±2.7) $\mu g/m^3$ for the NHS cohort and 18.3 (±3.1) $\mu g/m^3$ for the HPFS cohort. In the NHS cohort, of the total 74,412 female nurses who were followed up for 13 years, 3,784 developed T2DM. The IR was 4.5 per 1,000 person-years. After adjustment for potential confounders, the HR per 10 $\mu g/m^3$ increase in $PM_{2.5}$ was 1.05 (95% CI: 0.87, 1.22). In the HPFS cohort, of the total 15,048 male health professionals, 688 developed T2DM during the 13 years of follow-up. The IR was 4.0 per 1,000 person-years. After adjustment for potential confounders, the HR per 10 $\mu g/m^3$ increase in $PM_{2.5}$ was 1.18 (95% CI: 0.81, 1.71). The authors did not assess the potential for effect modification by covariates.[88]

The same authors published two cohort studies on the same cohort of African American women as part of the Black Women's Health Study (BWHS) situated in Los Angeles, USA, with differing follow-up times and methods of $PM_{2.5}$ modeling and metric assignment.[84,87] The women were aged 21 to 69 years at baseline who were subscribers to the Essence magazine. Both studies identified incident cases of T2DM based on a self-reported biennial mailed questionnaire and web-based health questionnaires, which were validated in a validation study with confirmed medical records provided by physicians. In the earlier study with a total of 10 years of follow-up, Coogan et al. (2012) identified 183 cases of T2DM out of 3,992 women. The authors used interpolation with kriging models and local monitoring data to quantify $PM_{2.5}$ as annual time-varying exposure to each participant, by linking residential postal code. The overall mean $PM_{2.5}$ was 20.7 (±2.1) $\mu g/m^3$. The IR was 5.4 per 1,000 person-years and the adjusted IRR per 10 $\mu g/m^3$ increase in $PM_{2.5}$ was 1.63 (95% CI: 0.78, 3.44).[87]

In the more recent BWHS study, Coogan et al. (2016) identified 4,387 cases of T2DM from 43,004 women who were followed up for 15 years, five years more than their earlier study. They used a hybrid modeling that incorporated LUR and BME to quantify $PM_{2.5}$ concentration and assigned $PM_{2.5}$

exposure as a time-fixed 9-year average, by linking participants to their residential postal codes. They were able to account for residential mobility. The overall mean $PM_{2.5}$ was 13.9 μg/m^{3}with an IQR of 2.9 μg/m^3. The IR was 9.7 per 1,000 person-years and the adjusted HR per 10 μg/m^3 increase in $PM_{2.5}$ was 1.52 (1.13, 2.06), presenting a statistically significant association that was not found in their 2012 study.[84]

2.2 Europe

There were four papers from Europe (specifically Denmark, Italy, and Germany) that examined the association between long-term exposure to $PM_{2.5}$ and the incidence of T2DM.[89–92]

Enrolling women aged 44 or older in 1993 from the Danish Nurse Cohort, Hansen et al. (2016) followed up 24,174 female participants free of T2DM at baseline until 2012 for an average of 15.3 years. T2DM data were collected from the National Diabetes Register. $PM_{2.5}$ concentrations were modeled using the high-resolution Danish air pollution dispersion modeling system (AirGIS), a GIS-based dispersion model. Concentrations were linked to each individual at residential address, and annual residential mobility was accounted for. Personal exposure was assigned as a 5-year moving average into the statistical model. The regional mean concentration of $PM_{2.5}$ was 18.1 (±2.8) μg/m^3 and the IQR was 3.1 μg/m^3. The mean age at baseline was 54.0 (±8.2) years. By the end of follow-up in 2012, 1,137 new cases of T2DM were identified with an IR of 3.0 per 1,000 person-years. The adjusted HR was 1.40 (95% CI: 1.03, 1.90) per 10 μg/m^3 increase in $PM_{2.5}$ exposure. Females who were never smokers, were obese, or had existing cardiovascular disease were at higher risk. Their findings additionally suggested a linear association between exposure to $PM_{2.5}$ and incidence of T2DM.[89]

Renzi et al. (2018) examined the association using participants of the Rome Longitudinal Study (RoLS) from Italy. Participants were aged 35 years or older and free of T2DM at baseline, and personal

data were obtained from the Regional Health Information System for a total of 1,319,193 individuals. T2DM was defined by administrative data codes for having qualified for free healthcare in the Exempt from Copays Registry, admitted to hospital with T2DM diagnosis, or prescribed hypoglycemic drugs. Although blood samples from physical examinations were not done in the total cohort, validation test with HbA1c of greater than or equal to 5.7% on a random sample was done. $PM_{2.5}$ concentrations were estimated from LUR that additionally included 20 sampling sites. Concentrations were linked to participants by residential address at baseline only. The regional mean concentration of $PM_{2.5}$ was 19.6 $\mu g/m^3$ with an IQR of 1.7 $\mu g/m^3$. Approximately 55% of participants were female. Over the period from 2008 to 2013, with a mean follow-up of 5.2 years, a total of 65,955 new cases of T2DM were identified. The IR was 9.6 per 1,000 person-years and 10.1 when standardized for age and gender. The adjusted HR was 0.99 (95% CI: 0.95, 1.04) per 10 $\mu g/m^3$ increase in $PM_{2.5}$, suggesting no association. It should be noted that, as it was based on an administrative dataset, information on BMI, physical activity, and diet, were not available, and these factors could profoundly confound the association.[90]

In the Heinz Nixdorf Recall Study (HNRS) conducted in Germany, Weinmayr et al. (2015) followed up participants who were aged 45 to 75 years at baseline for an average of 5.1 years of follow-up during the period from 2000 to 2008. Personal information was collected from self-administered questionnaires, face-to-face interviews, and a physical examination that included lab tests. T2DM was defined as self-reported physician diagnosis, taking diabetes medication (ATC code A10), RPG of 11.1 mmol/L or higher, or a FPG of 7.0 mmol/L or higher. $PM_{2.5}$ concentrations were estimated from the European Air Pollution Dispersion and Chemistry Transport Model (EURAD-CTM) at a 1×1 km spatial resolution and linked to each individual by residential address at baseline only. Exposure was assigned as a 2-year (2001 to 2002) time-independent average into the statistical model. The regional mean concentration of $PM_{2.5}$ was 16.7 $\mu g/m^3$ with an IQR of 2.3 $\mu g/m^3$. Of the 3,607 participants free of T2DM at baseline, 331 new cases were identified at the end of follow-up. The adjusted relative risk (RR) was

1.34 (0.59, 3.06) per 10 μg/m^3 increase in $PM_{2.5}$, indicating no significant association, which was not significantly modified by gender or BMI.[92]

More recently, Lucht et al. (2020) performed another analysis of the same cohort as Weinmayr et al. (2015), with a longer period of follow-up (10.3 years). Lucht et al. (2020) found that the adjusted RR per 10 μg/m^3 increase in $PM_{2.5}$ was 1.34 (95% CI: 0.58, 3.12), and the result was similar to that from the report by Weinmayr et al. (2015).[91]

2.3 Asia

Among four cohort studies conducted in Asia,[93–96] one covered mainland China, two looked at Taiwan, and one was done in Hong Kong. Liang et al. (2019) examined the association between T2DM incidence and long-term exposure to $PM_{2.5}$ in China using four large-scale cohorts from the Prediction for Atherosclerotic Cardiovascular Disease Risk in China Project (China-PAR Project), specifically the China Multi-Center Collaborative Study of Cardiovascular Epidemiology (ChinaMUCA) 1992–1994, ChinaMUCA 1998, the International Collaborative Study of Cardiovascular Disease in Asia (InterASIA), and the Community Intervention of Metabolic Syndrome in China & Chinese Family Health Study (CIMIC). Participants aged 35 to 74 years at baseline who were free of T2DM were included in the analytical cohort. The study was run from 2004 to 2015, with a follow-up of 11 years. T2DM incidence status was ascertained from a standardized questionnaire interview, physical examination, and blood sample. T2DM case was defined as having a diagnosis of diabetes, use of insulin or hypoglycemic agents, or a FPG greater than or equal to 7.0 mmol/L. Ground-level ambient concentrations of $PM_{2.5}$ were approximated from satellite-based estimates at a 10×10 km spatial resolution. Personal exposure was linked to participants based on residential postal code as a proxy for average long-term time-space activity, and assigned as a time-weighted average over the 11-year follow-up period into the statistical

model. Of the 88,397 participants, 6,439 developed T2DM over the 11-year follow-up, with an IR of 11.1 per 1,000 person-years. The mean age was 51.7 (±11.7) years at baseline, 60.2% were female, the mean BMI was 23.7 (±3.5) kg/m^2), and 26.0% were smokers. The regional mean $PM_{2.5}$ concentration was 79.1 (±13.8) μg/m^3. The adjusted HR was 1.16 (95% CI: 1.06, 1.26) per 10 μg/m^3 increase in $PM_{2.5}$. Effect modification, leading to higher risks in certain subgroups, was identified for young- to middle-aged subjects, females, rural subjects, subjects with a BMI less than 25 kg/m^2, non-smokers, and non-hypertensive subjects.[93]

Lao et al. (2019) carried out a longitudinal cohort study in Taiwan, based from a private firm, the MJ Health Management Institution, that required a paid membership. Their cohort included participants 18 years or older who were free of T2DM at baseline that had at least two medical visits to the firm. T2DM cases were determined from self-reported physician-diagnosis in a standard self-administered questionnaire or a FPG greater than or equal to 7 mmol/L from a medical examination. A participant's home address or business address was used to link satellite-derived $PM_{2.5}$ concentrations to each individual at a spatial resolution of 1 × 1 km, and residential mobility per visit interval (1.2 ±4.1 years between each visit) was accounted for. The satellite-based estimates were further calibrated using AERONET in Taipei, Taiwan, and validated with measures from 70 ground-level monitors in the region. Personal exposure was assigned as a 2-year moving average into the statistical model. The regional mean $PM_{2.5}$ concentration was 26.7 (±7.4) μg/m^3, with an IQR of 6.3 μg/m^3. Over the study period from 2001 to 2014, the mean follow-up length was 6.7 years, during which 4,781 out of 147,908 participants developed T2DM. The mean age of participants was 38.3 (±11.5) years with a roughly equal gender distribution (female: 49.9%). The IR was 3.5 per 1,000 person-years. Adjusted HRs were presented as comparisons between quartiles. The adjusted HR was 1.16 (95% CI: 1.07, 1.26) for participants in the uppermost quartile (4th quartile: ≥28.0 μg/m^3) compared to participants in the lowermost quartile (1st quartile: <21.7 μg/m^3). Interestingly, the adjusted HR suggested greater risks for participants in the 3rd

quartile (24.1–27.9 μg/m^3) and 2nd quartile (21.7–24.0 μg/m^3) than was seen in the 4th vs. 1st quartile; specifically, the adjusted HRs were 1.27 (95% CI: 1.17, 1.38) for third and 1.18 (95% CI: 1.18, 1.39) for 2nd vs. 1st quartile, respectively. The association was stronger in participants with a BMI less than 23 kg/m^2 or in those who were alcohol drinkers.[94]

Li et al. (2019) utilized the Taiwanese National Health Insurance Research Database to recruit participants 20 years or older in Taiwan in 2001 to 2012. T2DM incidence was attained from the Longitudinal Health Insurance Database. T2DM was defined as an ICD-9 code 250 more than three times in a year, or the concurrent use of anti-hyperglycemic medication. Ground-level monitoring data from the EPA of Taiwan from 2006 to 2012, in addition to LUR, was used to determine $PM_{2.5}$ concentration. $PM_{2.5}$ data were assigned as a 7-year time-invariant average (2006 to 2012) and linked to each individual by place of most frequently visited medical facility at baseline. The regional mean concentration of $PM_{2.5}$ was 27.92 μg/m^3, ranging from 13.81 to 39.94 μg/m^3. The mean age of participants was 42.6 (±15.8) years, and there was roughly an equal split in gender (female: 51.3%). Over the median 12-year follow-up period, 48,611 of the 505,151 participants free of T2DM at baseline developed the disease. The 12-year cumulative incidence was 8.4%. The final HR, adjusted for all individual-level covariates, comorbidities, and regional covariates, was 1.11 (95% CI: 1.08, 1.13) per 10 μg/m^3 increase in $PM_{2.5}$. Gender, age, and hyperlipidemia demonstrated to be effect modifiers of this association, and the association was stronger in males, those aged 65 years or older, and those with hyperlipidemia.[95]

Li et al. (2019) additionally examined the association as comparisons between quartiles, just as was done by Lao et al. (2019). Using the 1st quartile (<27.01 μg/m^3) as a reference category, Li et al. (2019) found the adjusted HR increased with increasing quartiles; specifically, the adjusted HR was 1.01 (95% CI: 0.99, 1.04) for the 2nd quartile (27.01–34.15 μg/m^3), 1.06 (95% CI: 1.03, 1.09) for the 3rd quartile (34.16–39.03 μg/m^3) and 1.15 (95% CI: 1.12, 1.18) for the 4th quartile (39.04–47.20 μg/m^3) compared

with the 1st quartile suggesting an almost linear relationship between exposure to $PM_{2.5}$ and the incidence of T2DM.[95]

In comparison, Lao et al. (2019) found that the concentration-response association between exposure to $PM_{2.5}$ and the incidence of T2DM was non-linear, in which with increasing exposure came increasing risk, but decreased in the uppermost concentrations. However, the authors cautioned interpretation as the phenomenon may have been due to heterogeneities in the populations as well as unidentified confounders.[94]

Qiu et al. (2018) performed a cohort study in Hong Kong, recruiting participants from the Elderly Health Services Cohort on a volunteer basis, who were aged 65 years or older at baseline. New cases of T2DM were determined from hospital discharge records (ICD-9 codes 250). The study began in 1998 and participants were followed up until 2010 with an average of 9.8 years. Of the 53,905 participants, 806 developed T2DM during follow-up. The IR was 1.5 per 1,000 person-years,[96] which is relatively lower compared to the other Asian cohorts.[93–95] Satellite-based $PM_{2.5}$ concentrations were used to estimate ground-level concentrations at a 1×1 km spatial resolution. Personal exposure was linked using residential postal codes and annual residential mobility was accounted for. Exposure was assigned as an annual time-varying exposure into the statistical model. The overall regional mean $PM_{2.5}$ concentration was 35.8 $\mu g/m^3$, with an IQR of 3.2 $\mu g/m^3$. The mean age was 72 years and 65.9% were female. The adjusted HR was 1.55 (95% CI: 1.18, 2.03) per 10 $\mu g/m^3$ increase in $PM_{2.5}$. Effect modification by gender suggested that the association was stronger in females. A similar finding was found by Liang et al. (2019), but an opposing finding by Li et al. (2019), suggesting more research on this potential effect modifier is needed. Sensitivity analysis, in which exposure was assigned as time-independent 3-year average at baseline, resulted in similar results. Furthermore, the concentration-response association between exposure to $PM_{2.5}$ and the incidence of T2DM was in linear fashion, similar to what Li et al. (2019) found in their study, but dissimilar to what was found in the study by Lao et al. (2019).[96]

2.4 Gaps in the Current Literature

There were a total of 21 cohort studies that examined the relationship between long-term exposure to $PM_{2.5}$ and the development of T2DM. Thirteen of the cohorts were taken from North American countries, four were from European countries, and the remaining four from Asian countries. The minimum age limit for eligible inclusion in these studies ranged from 18 to 65 years of age. Although a few cohort studies were gender-specific, the majority of cohorts included analysis on both genders. The maximum follow-up time considered in these studies ranged from 5 to 22 years. The method of exposure assessment ranged greatly, such as LUR, fixed-site monitoring, and satellite imaging. The majority of studies used incident T2DM as the measurable outcome. Nine studies assigned $PM_{2.5}$ exposure as a time-fixed variable in their statistical model, and 12 studies assigned $PM_{2.5}$ exposure as a time-varying variable in the main analysis or sensitivity analysis. The annual mean concentration of $PM_{2.5}$ in each study ranged from as low as 4.1 to as high as 79.1 $\mu g/m^3$, and the WHO air quality guidelines suggest a cut-off of 10.0 $\mu g/m^3$.[35]

Of the results, all but one study were suggestive that increased exposure to $PM_{2.5}$ was associated with development of or death due to T2DM, but only 14 showed statistical significance, suggesting the need for more replication in more populations. Notably, the findings of this literature search suggest a few important gaps in research. Firstly, only four studies were conducted in Asia, and none focused on rural populations. It is known that Asians have differing risk profiles for diabetes compared to their Western counterparts.[97,98] The annual mean concentrations of $PM_{2.5}$ were not comparable to other regions, as all except one study were done on regions with annual mean concentrations of $PM_{2.5}$ less than 40 $\mu g/m^3$. Furthermore, the source of $PM_{2.5}$ has been shown to differ between regions,[32] lending to the issues with generalizing such findings to our population of interest.

In addition limited generalizability, there has been inconclusive data on effect modification and the individual roles of specific constituents of PM2.5. All but one study assessed effect modification on a multiplicative scale rather than an additive scale, which has been suggested to better suit effect modification that is hypothesized to be biological in nature.[99,100] Air pollution epidemiologic studies commonly assess the relationship between health and PM2.5 mass, assuming that the effects of all particles meeting size requirements elicit the same toxic effects. However, constituents of PM2.5 mass do have differing toxicities, and may have differing effects on health.[101] Hence, the present study serves to fill the gap in researched populations that is currently limiting the epidemiologic research on the relationship between long-term PM2.5 exposure and the risk of T2DM.

CHAPTER 3. Association between long-term exposure to $PM_{2.5}$ and incidence of T2DM in a rural cohort in China (Article 1)

Abstract for Article 1

Background: Type 2 diabetes mellitus (T2DM) is an increasingly prevalent disease worldwide, with drastic increases in low- and middle-income countries. China is one of most severely polluted countries, with very high concentrations of air pollutants such as fine particulate matter ($PM_{2.5}$). The health effects of long-term exposure to $PM_{2.5}$ on respiratory diseases have been well defined, but only recently have researchers begun to examine the relationship of $PM_{2.5}$ and T2DM. Most studies are done in developed countries that have many years of monitoring data, allowing for long-term effects to be readily researched. Furthermore, of the limited research done in China, most took place in urban megacities.

Objective: The main purpose of this article is to investigate whether long-term exposure to $PM_{2.5}$ is associated with increased incidence of T2DM in a rural area of China. The second purpose is to examine the robustness of our main findings to different temporal metric assignments of $PM_{2.5}$ exposure commonly used in the literature.

Methods: This prospective cohort study was conducted from 2006–2014 with follow-up until December 31, 2018. The cohort consists of residents from rural townships and villages from the Deqing County, Zhejiang Province, China, who were 18–64 years old and free of T2DM at baseline. Personal exposures to $PM_{2.5}$ were determined by linking satellite-derived estimates to participants' residential addresses at baseline. Cases during follow-up were defined from results of fasting plasma glucose (FPG) tests, self-reported physician-diagnosis, or antidiabetic medication use, and/or ascertained from electronic health records. Cox proportional hazard models were used to calculate hazard ratios (HR) and 95% confidence intervals (CI). Potential confounders, chosen *a priori*, such as exercise, obesity, and family history of diabetes, were adjusted for in the models. Potential effect modification was assessed by including multiplicative interaction terms into the models.

Results: Greater long-term exposure to $PM_{2.5}$ was associated with increased risk of T2DM. In the model adjusted parsimoniously for age, body mass index (BMI), and education, the HR per 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure was 1.22 (95% CI: 1.00, 1.49). In the model adjusted fully for all individual-level covariates, the HR per 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure was 1.26 (95% CI: 1.02, 1.54). There was evidence of effect modification by smoking status, with increased risk in smokers vs. to non-smokers.

Conclusions: Our findings suggest that long-term exposure to $PM_{2.5}$ is associated with incidence of T2DM in rural adults. Developing more high quality cohort studies, with long follow-up periods and implemented in locations of differing pollution levels, will benefit our understanding of this relationship.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AOD	Aerosol optical depth
AQG	Air Quality Guidelines
BC	Black carbon
BMI	Body mass index
CHD	Coronary heart disease
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
FPG	Fasting plasma glucose
GEOS-CHEM	Goddard Earth Observing System Chemical Model
HR	Hazard ratio
IQR	Interquartile range
MAIAC	Multi-Angle Implementation of Atmospheric Correction
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NH_4^+	Ammonium
NO_3^-	Nitrate
OC	Organic carbon
$PM_{2.5}$	Fine particulate matter
PM_{10}	Coarse particulate matter
SD	Standard deviation
SeaWIFS	Sea-Viewing Wide Field-of-View Sensor
SO_4^{2-}	Sulfate
T2DM	Type 2 diabetes mellitus
WHO	World Health Organization

3.1 Introduction

Diabetes mellitus is an increasingly prevalent disease worldwide. The World Health Organization (WHO) reported that approximately 9% of the global population was affected by diabetes, which was responsible for 2.7% of all global deaths.[1] Over 99% of all cases of diabetes are type 2 diabetes mellitus (T2DM).[2] By 2035, it is estimated that the number of people affected by diabetes will have increased to a staggering 592 million people, with the sharpest increases expected in low- and middle-income countries, such as China.[3] China has experienced an increased prevalence of diabetes from 5.5% to 11.6% in the past decade alone. Most notably, the rate at which diabetes prevalence has increased in China far surpasses that of many other low- and middle-income countries worldwide. China's rapidly increasing prevalence is much higher in comparison to the rest of the world.[1,4]

Ambient air pollution requires our immediate attention; 4.2 million deaths in 2015 were estimated to be attributable to ambient air pollution with nearly half of those deaths (2.2 million) occurring in the western Pacific, a region that includes China. Future trend analyses predict that by 2050, the number of deaths attributable to ambient air pollution will have increased 50%.[5] Low- and middle-income countries have made major economic growth over the past several decades that are supported by rapid development of industries and services.[6] This expansion, accompanied by increases in energy consumption, reflect a worrisome increase in ambient air pollution.[5] Of particular interest is fine particulate matter ($PM_{2.5}$); particulate matter is a complex mixture of small particles of numerous components such as organic chemicals, metals, and soil or dust particles. The size of the particle determines the route of entry and its potential to cause adverse health effects. Whereas coarse particulate matter (PM_{10}), with an aerodynamic diameter of less than 10 μm, enters through the throat and nose, due to its small size, $PM_{2.5}$, which has an aerodynamic diameter of less than 2.5 μm, can pass even further into the alveolar region of the lung.[7] Under their air quality guidelines, WHO defines the

standard for annual $PM_{2.5}$ concentrations to be less than 10 μg/m^3.[8] Not only do most prefectures in China exceed this value, but they can also reach above 35 μg/m^3, concentrations at which human illness rates increase immensely.[9,10] Researchers have been interested in assessing whether long-term exposure over several years leads to amplified risks above acute risks, hypothesizing that long-term exposure promotes or enhances the progression of chronic disease states.[11] The health effects of $PM_{2.5}$ on cardiovascular and respiratory diseases are well defined;[5] however, only recently there have been findings on the association between ambient air pollution and T2DM with inconsistent results.[12,13]

With the rising levels of air pollution concentrations and prevalence of T2DM worldwide, a growing number of studies have been published in attempt to explore and explain the association between long-term exposure to $PM_{2.5}$ and T2DM. To demonstrate an etiologic relationship, prospective observational studies are needed; however, very few studies have been made in China. More than 90% of all studies on this association have been examined in developed countries.[4,14] The results are ungeneralizable to Asian populations, particularly because both diabetes prevalence and air pollution concentrations are reported to be much higher amongst developing Asian countries.[3,15] Furthermore, evidence suggests that risk profiles for diabetes of East Asian populations differ greatly from their North American and European counterparts.[15] Therefore, it is crucial that further exploration into the association between exposure to $PM_{2.5}$ and incidence of diabetes are explored in China and other developing countries. Secondly, most studies performed in China settings focus on urban areas. It is known that $PM_{2.5}$ concentrations are not comparable between urban and rural areas.[9] Although the focus on urban cities is understandable given that megacities generate a large amount of air pollutant emissions,[6] rural cities should not be ignored, especially considering the expansion of medium-sized cities and small towns around these larger cities.

The 2017 Lancet Commission report on pollution and health outlined the need for pollution-related research to "explore emerging causal links between pollution [and] disease" and to "identify and map pollution exposures particularly in low-income and middle-income countries".[5] The proposed thesis aims to address these two gaps in the current research. Advancements in technologies to map pollution, such as satellite imaging, has allowed researchers to examine associations between air pollution and a wider range of diseases than historically possible. The effect of ambient air pollution on T2DM is still new and has yet to be fully characterized. Furthermore, the association between ambient air pollution and T2DM have not been well studied outside of the mega-cities such as Beijing, which attract most of the attention.[5]

In this work, the association of T2DM incidence and long-term exposure to $PM_{2.5}$ was examined using data from a rural cohort in Deqing County, Zhejiang province, China. The aim of this work is to confirm whether greater long-term exposure to $PM_{2.5}$ is associated with increased risk of developing T2DM in a rural Chinese population, and secondly, to examine whether the results of our main analysis are sensitive to implementation of different temporal metrics commonly used in the literature.

3.2 Methods

Study Design and Population

The Rural Deqing Cohort Study was first conducted as a cross-sectional study in 2006 to 2007,[16] and serves as the present study's baseline. Additional study participant recruitment and in-person interviews were conducted in 2008, 2011, 2012, 2013, and 2014. Residents were randomly cluster sampled from a total of eight rural townships, including 65 villages in Deqing County, Zhejiang Province, China.[17] Study participants were followed up every two to four years.[18] The present study was

conducted over the period 2006 to 2014 and participants were followed up until December 31, 2018 in this cohort study.

The target population included adults who were 18 to 64 years old at study entry and resided in one of the sampled rural townships in Deqing County. Adults without local residency (i.e. temporary workers or university students not living in the county during the study period) were excluded.[16] To be eligible for study enrolment, local adult residents had to have no plans to leave Deqing County, whether for change in work or home location. Furthermore, eligible participants had to indicate agreement to participate in the study by signing a written informed consent form, answering interview questions from the baseline survey questionnaire, and agreeing to attend follow-up physical examinations.[17]

Local authorities provided a list of names and household addresses to identify potentially eligible adult residents in the rural townships. Local health workers, trained by academic investigators to act as data collectors, visited these households and held in-person interviews to confirm eligibility and recruit willing individuals. Alternatively, interviews were held at nearby physical examination centers if potential participants were not available at the time of the data collector's home visit. If successfully recruited and informed consent was attained, the data collector administered a baseline survey questionnaire to collect information on demographics, lifestyle characteristics, and disease history that was planned to take approximately 15 minutes. Study participants were asked to undergo a physical examination by a physician, during which anthropometric measurements were taken and a 5-ml blood sample was drawn.[16]

Participation at baseline and at follow-up was completely voluntary. Participants received free medical check-ups regularly and were provided advice and additional information on health and diabetes prevention as a result of participating in the study. Furthermore, participants received a living allowance as compensation for their time.

Exposure Assessment

Ground-level $PM_{2.5}$ concentrations were approximated from satellite-derived ambient $PM_{2.5}$ estimates.[19] In brief, this dataset combines satellite retrievals and simulation of aerosol optical depth (AOD) from multiple sources, specifically Multi-angle Imaging SpectroRadiometer (MISR), Moderate Resolution Imaging Spectroradiometer (MODIS) Dark Target, MODIS and Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) Deep Blue, MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC), and Goddard Earth Observing System Chemical Model (GEOS-Chem) with the geophysically-based simulated relationship between aerosol optical depth and near-surface $PM_{2.5}$ concentrations. GEOS-Chem simulation, a chemical transport model, is used to partition $PM_{2.5}$ mass into composition, namely OC, BC, SO_4^{2-}, NO_3^-, NH_4^+, and soil dust. Ground-based observations from the recently expanded monitoring network over mainland China are then incorporated using monthly geographically weighted regressions at a 0.01° × 0.01° resolution from 2014 to 2016. The resultant annual average $PM_{2.5}$ concentrations were consistent with out of sample, cross-validation observations (R^2 = 0.78). $PM_{2.5}$ concentrations prior to this time period are based on the relative changes of Atmospheric Composition Analysis Group's long-term product.[19–21]

Annual average ground-level $PM_{2.5}$ concentrations based on satellite-derived estimates were available at a 1 × 1 km spatial resolution for linkage in Deqing County. Annual average concentrations were linked to each study participant based on spatial and temporal differences. Firstly, $PM_{2.5}$ concentrations at the 1 × 1 km grid of land that includes a participant's home residence, served as the spatial linkage. Secondly, the annual average $PM_{2.5}$ concentrations in the years from study entry to development of T2DM or loss-to-follow, served as the temporal linkage. These two factors in combination act as a surrogate measure of true long-term personal exposure to outdoor ambient $PM_{2.5}$. A key assumption, therefore, is that concentrations at place of residence and annual averages are an

adequate approximation of theoretical true exposure based on personal time-activity and microenvironment patterns. Participants with missing information on air pollution exposure were excluded from the analytical dataset. Details on how exposure metrics were assigned are presented in **Supplementary Material 1** (**Table S1.1–S1.8**).

Outcome Ascertainment

Prevalent cases of T2DM were identified from self-reported diagnosis from responses in the interview's baseline questionnaire. Specifically, participants were asked "Has a doctor ever diagnosed you with diabetes?" with multiple choice responses 'yes', 'no', and 'do not know', and a follow-up question "If yes, confirmed time ___ years, and is ___ type." Additional questions gave further context, specifically: "Have you ever measured blood sugar levels in a hospital?" with multiple choice responses 'yes', 'no', and 'do not know' and a follow-up question "If yes, did the doctor say your blood glucose was high, normal, low, or do not know?"

Incident cases of T2DM were identified over the follow-up period from the results of blood tests or self-reports. Specifically, T2DM diagnosis was defined as having fasting plasma glucose (FPG) of 7.0 mmol/L or higher, self-reported physician-diagnosed diabetes, or use of antidiabetic medication. T2DM defined with FPG levels was determined from blood samples collected during the physical examination after an overnight fast of at least 8 hours. A FPG test was performed within 2 hours of blood sample extraction using the glucose oxidase method.[16] T2DM defined with self-reported physician-diagnosed diabetes or use of antidiabetic medication was determined from responses in the interview's baseline questionnaire.

T2DM was coded as a dichotomous outcome ('yes' or 'no') and participants with missing information on T2DM were excluded from the analytical dataset. T2DM status was ascertained from electronic health records of Deqing County up until December 31, 2018. For the majority of participants, electronic health records were available; electronic health records were missing for 1,899 (6.7%) participants.[17]

Other Covariates

All additional participant characteristics were determined from information collected in the survey questionnaire, other than anthropometric measurements (such as weight and height) from the physical examination. Except for missing data on T2DM, $PM_{2.5}$, and smoking status, missing covariate responses were coded as their own category and thus retained in the dataset. The purpose of retaining missing covariate responses as their own category is to prevent excessive sample size loss, which would reduce power.[22]

Gender was coded as a dichotomous covariate ('male' or 'female'). Participants missing information on gender were excluded from the analytical dataset.

Age (years) at baseline was calculated from date of birth to date the survey was administered and was rounded down to the nearest whole number age; for example, 44.2 and 44.8 years old at baseline would both be rounded down to 44 years old in the analytical dataset. Discrete continuous age was then recoded as a categorical covariate, as age groups at 10-year intervals (specifically, '18–39', '40–49', '50–59', '60–69', and '≥70'). Participants missing information to determine their age at baseline were included in the analytical dataset and coded as an additional category ('unknown').

Family history of diabetes was determined from the results of linking questionnaire responses. Family members were linked based on the listed family members identified in the questionnaire. Listed family members were identified as either living locally or not, and local members were asked to fill the questionnaire and indicate T2DM status. In addition to the names given during the interview, participants' family members could be found linking 'family ID'. Family history of diabetes was coded as a dichotomous covariate ('no' or 'yes') in the analytical dataset. Family history of diabetes was defined as 'yes' if they had at least one first-degree relative with diabetes.[17] Participants with missing information on family history of diabetes were included in the analytical dataset and coded as an additional category ('unknown').

History of chronic disease(s) was determined from responses to one multiple choice question on the questionnaire. Specifically, participants were asked "Did you ever have a doctor diagnose you with any of the follow illnesses?" with a list of various illnesses given as multiple choice responses, as well as an option to indicate 'other diseases ___', 'none', and 'do not know'. Of the illnesses, responses considered for history of chronic disease(s) include: hypertension, asthma, hemorrhagic stroke, ischemic stroke, coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD), and tumours. History of chronic disease(s) was coded as a dichotomous covariate ('no' or 'yes') and was defined as 'yes' if the participant had at least one chronic disease (non-specific) and 'no' for all other possibilities, including participants who responded 'no', 'do not know', or were missing responses.

Participants were considered smokers if they continually smoked for greater than six months and at least one cigarette daily.[16] Participants with missing information on smoking status were excluded from the analytical dataset. Smoking status was coded as 'never smoker' or 'ever or current smoker'.

Passive smoking exposure was defined as the response to the question "Do you often have people around you who smoke?" Prompts included questions about whether a person at home or a

friend smoke around them. Participants were coded as a dichotomous covariate ('no' or 'yes'). Participants missing information on passive smoking exposure were included in the analytical dataset and coded as an additional category ('unknown').

Vegetable consumption was measured as the approximate amount participants' ate per week on average last year and was a categorical variable ('<1.0', '1.0–1.9', and '≥2.0' kg/week). Participants missing information on vegetable consumption were included in the analytical dataset and coded as an additional category ('unknown').

Similarly, fruit consumption was measured as the approximate amount eaten per week last year. Fruit consumption was categorized as '<0.5', '0.5–1.4', and '≥1.5' kg/week. Participants with missing information on fruit consumption were coded as an additional category ('unknown').

Exercise was defined by the response to the question, "In the last year, did you exercise?" Subsequent exercise-related questions included frequency per week, duration of one session, and preferred training method, such as running, brisk walking, play ball, martial arts, and dancing. For the purposes of the present analysis, exercise was coded as a dichotomous covariate ('no' or 'yes'). Participants with missing information on exercise were coded as an additional category ('unknown').

Body mass index (BMI) was measured during the physical examination by a doctor and was calculated based on the participant's exact height (cm) and weight (kg). BMI was categorized into normal or underweight (<25.0 kg/m^2), overweight (25.0–29.9 kg/m^2), or obese (≥30 kg/m^2), following WHO international guidelines.[23]

Education, dichotomized as 'junior high school or higher', 'no junior high school diploma', or 'unknown' for participants with missing information on education, was taken from the question on education status with the following multiple choice options: 'illiteracy', 'not graduated from elementary

school', graduation from elementary school', 'graduation from junior high school', 'graduation from high school (including technical secondary school and vocational high school)', or 'college degree or above'. Graduation from junior high school was chosen as the cut-off to dichotomize the education covariate for the analytical dataset because education is compulsory in China from elementary school to junior high school (the Grade 9 equivalent in Canada).[24]

Occupation was included in the analytical dataset as a dichotomous variable, 'farmers' or 'other occupations', where all other occupations included occupations from the multiple choice answers: 'teachers', 'managers', 'officers', 'staffers', 'drivers', and 'students'. Occupation was dichotomized with farmer as the sole major category, due to the large percentage of farmers in rural China, as well as their differing outdoor air pollution exposure.

Socioeconomic status was included as categories of 'low', 'medium', or 'high', based on self-reported family income. The questionnaire asked what the family's perceived income level was, relative to other families living in the same township and is representative of the household.[16] Categories in the questionnaire included: 'low', 'medium', 'high', and 'wealthy', with 'high' and 'wealthy' being combined into one category for inclusion of the family income covariate into the analytical dataset.

3.3 Statistical Analysis

Main Analysis

Annual average concentrations of $PM_{2.5}$ over the study period was presented in a graph and the distribution of personal $PM_{2.5}$ exposure for cohort participants as a histogram, to assess whether the concentrations varied temporally and the exposures amongst participants was normally distributed, respectively.[25] Baseline characteristics for the total cohort overall was also presented according to

diabetes case status at the end of follow-up. The characteristics were presented as frequencies (percent) for categorical variables and as means [standard deviation (SD)] for continuous variables, assuming normal distribution. Baseline characteristics for the total cohort were additionally presented according to quartiles of baseline $PM_{2.5}$. The purpose of presenting baseline characteristics according to outcome and exposure is to identify covariates that appear associated with both and therefore having higher potential to be a confounder of this association of interest.[26] Person-years of follow-up were calculated from the date of entry into the study to the date of T2DM diagnosis, death, last year of contact, or to the end of the study, whichever came first.

The number of events and incidence rates of T2DM for every baseline characteristic and quartiles of $PM_{2.5}$ exposure were presented to assess whether incidence rates varied by baseline characteristic or $PM_{2.5}$ quartile categories.

Cox proportional hazard regression models was implemented to examine the association between long-term exposure to $PM_{2.5}$ and the incidence of T2DM and is a commonly used model in studies with similar purposes.[27–29] Using the Cox proportional hazards models, hazard ratios (HR) and 95% confidence intervals (CI) were estimated for a 10-$\mu g/m^3$ increase in $PM_{2.5}$, an increment that is consistent and comparable with most of the current literature.[30,31] HRs were additionally reported for incidence of T2DM for participants compartmentalized into groups and according to $PM_{2.5}$ quartiles to assess whether the assumption of linearity was feasible.[32–35]

Satellite-derived annual ambient $PM_{2.5}$ concentrations were assigned to each individual and approximated personal exposure in three different ways: baseline, 3-year average at baseline, and yearly time-varying exposure.[28,36,37] Since baseline and 3-year average are time-fixed assignments of $PM_{2.5}$ exposure, each participant has one row of observations. In comparison, yearly time-varying exposure requires the counting process data set-up, and followed guidelines by Andersen & Gill.[27,38]

Data set-up for the counting process requires the creation of multiple observations for each participant, where each observation represents one person-year of time-to-event follow-up.[39,40]

Potential confounding was accounted for by adjusting for covariates in the regression model. Covariates that were hypothesized to be potential confounders were chosen *a priori* and identified based on the current literature.[41,42] Three levels of confounder adjustment were generated: unadjusted model, parsimoniously-adjusted model, and fully-adjusted model. The parsimoniously-adjusted model includes only covariates that showed evidence of being confounders based on statistical analysis; in other words, each covariate was included into the base model individually, and the percent change of the β coefficient for $PM_{2.5}$ was calculated. A percent change greater than or equal to 10% was the chosen cut-off for an indication of confounding. All covariates that led to a percent change greater than or equal to 10% were then added into the parsimoniously-adjusted model. The parsimoniously-adjusted model was created to present results with less potential for overadjustment, which can lead to inflation of variance and loss of power.[22,42,43] In comparison, the fully-adjusted model includes all covariates that were *a priori* hypothesized to be potential confounders, regardless of the test of percent change.[44–46]

Research on this relationship has shown that some covariates may modify the effect of $PM_{2.5}$ on the development of T2DM and should be examined as potential effect modifiers.[25,31,34,47–57] Effect modification was conducted on multiplicative scale by including interaction terms and generating a p-value for the interaction, for which a p-value for the interaction less than 0.05 was considered significant and an indication of effect modification. If no interaction term presented a p-value less than 0.05, a p-value for the interaction less than 0.10 was considered significant, as is commonly utilized as a conservative method.[58] Inclusion of two effect modifiers into the same model was assessed by looking at the likelihood ratio tests comparing models with one effect modifier to models with two effect modifiers. A p-value for the likelihood test of less than 0.05 was considered significant.

All data management and statistical analysis were completed in SAS 9.4.

Sensitivity Analysis

Sensitivity of results to varying degrees of confounder adjustment was assessed by conducting models that were unadjusted, parsimoniously-adjusted, and fully-adjusted for all predetermined potential confounding covariates.[22,42,43]

Cohort studies on exposure to any air pollution and incidence of disease use different methods of assigning exposure in the statistical model. In the main analysis, $PM_{2.5}$ exposure was assigned as a continuous variable into the regression model and assumed to be approximated by ambient concentrations at baseline. Sensitivity of the results to differing choices in temporality of exposure assignment was assessed by conducting the analysis using additional proxy measures: full period time-fixed average (2006 to 2014), 3-year moving (time-varying) average, 3-year moving average with 1-year lag, yearly time-varying exposure with 1-year lag. A moving average is calculated as the average of years preceding and including the year of interest. For example, for the year 2011, a participant's exposure would be the average of concentrations of $PM_{2.5}$ at the place of residence in the years 2009, 2010, and 2011. Additionally, a 3-year moving average with 1-year lag for the year 2011 would be the average of concentrations of $PM_{2.5}$ in the years 2008, 2009, 2010. Detailed descriptions of assignment of the exposure metrics can be found in the **Supplementary Material 1** (**Tables S1.1–S1.8**).

Further analyses were performed to assess sensitivity to differing cohort definitions. Firstly, all participants that developed T2DM in their first year of entry were excluded, and the association between T2DM and long-term exposure to $PM_{2.5}$ was re-assessed. The purpose of this first subset cohort is to remove participants with underlying risk of T2DM independent from $PM_{2.5}$ exposure levels.[54,59] In

the second subset cohort, participants who contributed few person-years of follow-up (≤2) were excluded the association was reassessed. The purpose of this second subset cohort is to only include participants for which long-term exposure is more adequately and strictly defined, which can only be assumed based on the participants' total years of follow-up.

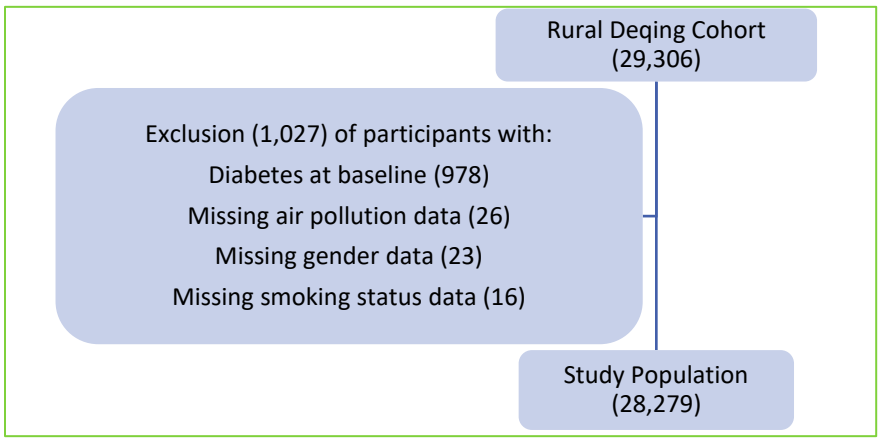
One major assumption made for the main analysis is the assumption of proportional hazards. The assumption of proportional hazards was assessed by categorizing the continuous variable $PM_{2.5}$ into quartiles and plotting the survivor function estimates against time, the log of negative log of survival against time, and the log of negative log of survival against log of time. Further tests include a model-based method for an interaction term between $PM_{2.5}$ and time as well as an <ASSESS> statement in SAS.

3.4 Results

Descriptive Statistics

The current study included 28,279 participants (**Figure 1**). Of the original 29,306 participants of the Deqing Rural Cohort, 1,027 (3.5%) were excluded from the current study. A total of 978 participants were excluded for having T2DM at baseline. The remaining exclusions were due to missing data on covariates deemed essential; specifically, 26, 23, and 16 participants were excluded for missing data on air pollution, gender, and smoking status, respectively.

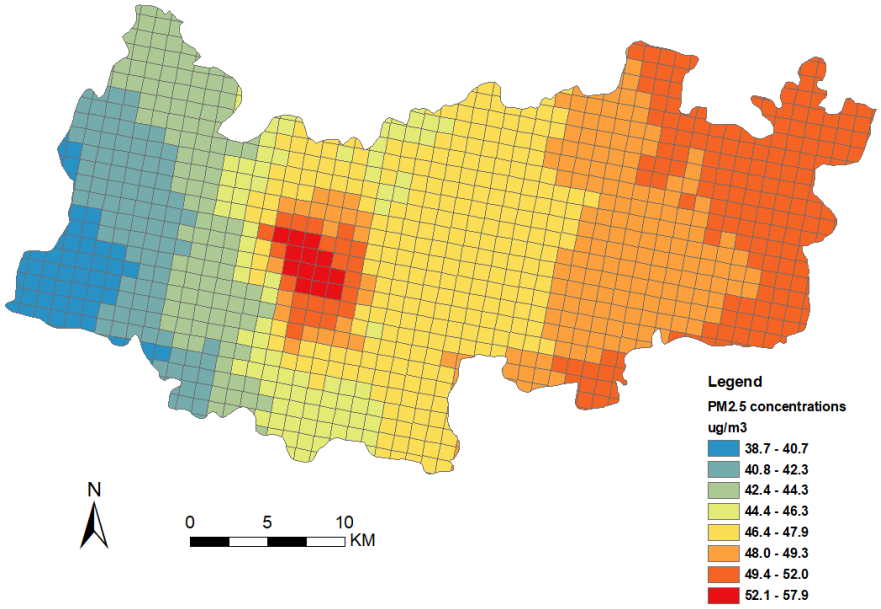
Figure 1. Flowchart of study participant inclusion and exclusion



Deqing County is situated in the North of Zhejiang, a coastal province in the West of Yangtze Delta, between latitude 30°261'–30°421'N and longitude 119°451'–120°211'E.[60] Deqing County has an area of approximately 938 km^2 with a total of 12 towns, including 166 villages, under its jurisdiction. The terrain is varied; the western portion of Deqing County is predominantly mountains, plains in the eastern portion, and hills in the central portion.[61,62]

The average annual $PM_{2.5}$ concentration over the study region in the baseline year is presented in a heat map of Deqing County on a 1×1 km grid (**Figure 2**). The regional distribution of $PM_{2.5}$ exhibits highest concentrations shown in red, in Wukang Town, the political, economic, and culture center of Deqing County.[62]

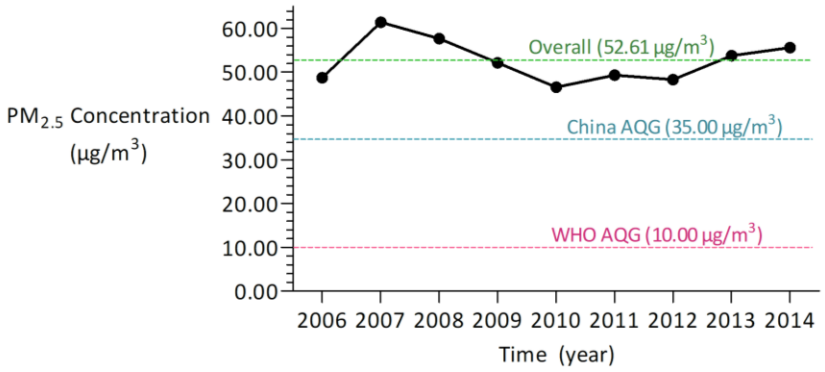
Figure 2. Heat Map of Spatial Variability of Ambient Fine Particulate Matter Concentration in Deqing County, in 2006



Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$.

Averaged over the period between 2006 and 2014, the overall temporal average $PM_{2.5}$ concentration was 52.61 $\mu g/m^3$ (**Figure 3; Table 1**). The individual $PM_{2.5}$ exposures for each participant were roughly normally distributed. The temporal trends in average annual $PM_{2.5}$ concentration remained consistently above the Air Quality Guidelines (AQG) set by Chinese State Council for China standards (35.00 $\mu g/m^3$) and the AQG set by WHO for international standards (10.00 $\mu g/m^3$).[63] Year-by-year, the average annual $PM_{2.5}$ concentration did not vary drastically (**Figure 3; Table 1**), suggesting that $PM_{2.5}$ concentrations assigned in the baseline year may serve as a sufficient proxy measure for true personal exposure.[25,36,37,59,64,65]

Figure 3. Temporal Trends in Overall Annual Ambient Fine Particulate Matter Concentration in Deqing County, from 2006 to 2014



Abbreviations: AQG, Air Quality Guidelines; $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$.

Table 1. Temporal Distribution of Annual Ambient Fine Particulate Matter Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	52.61	3.53	43.86	50.94	53.29	55.17	62.54	4.22
By year								
2006	48.74	6.63	39.42	46.83	48.58	49.50	82.92	2.67
2007	61.41	5.34	42.17	59.08	61.92	63.17	84.00	4.08
2008	57.66	3.53	48.92	56.75	58.75	60.17	67.58	3.42
2009	52.16	3.77	42.67	49.25	53.00	55.67	61.17	6.42
2010	46.53	3.00	40.33	45.08	46.58	49.25	56.92	4.17
2011	49.34	3.27	41.33	47.00	49.67	52.50	60.08	5.50
2012	48.31	2.81	41.33	46.50	49.08	50.58	59.00	4.08
2013	53.76	4.58	41.50	51.42	55.17	57.83	63.25	6.42
2014	55.59	3.48	45.67	55.25	56.83	57.83	67.17	2.58

Abbreviations: SD, standard deviation; IQR, interquartile range; $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$.

Over the follow-up period, 386 new cases of T2DM were identified (**Table 2**). At baseline, the cohort consisted of slightly more females than males (55.7% vs. 44.3%), very few participants had a family history of T2DM (1.4%), and less than a third had a chronic disease at baseline (30.0%). The majority of participants were farmers (73.1%) and did not complete a junior high school diploma (78.6%). Other details on smoking status, passive smoking exposure, vegetable consumption, fruit

consumption, exercise, BMI, and family income are presented in **Table 2**. Overall, the characteristics at baseline of the cohort are comparable to those of other cohort studies in China.[66–68]

Table 2. Baseline Characteristics of Study Participants, According to Type 2 Diabetes Mellitus Status at End of Follow-Up

Baseline characteristics	Total cohort		New cases of T2DM		Participants without T2DM at end of follow-up	
	No.	%	No.	%	No.	%
Overall	28,279	100.0	386	100.0	27,893	100.0
Gender						
Female	15,766	55.7	242	62.7	15,524	55.7
Male	12,513	44.3	144	37.3	12,369	44.3
Age at baseline (years)						
18–39	3,372	11.9	8	2.1	3,364	12.1
40–49	5,628	19.9	63	16.3	5,565	20.0
50–59	7,225	25.6	143	37.1	7,082	25.4
60–69	6,866	24.3	108	28.0	6,758	24.2
≥70	5,188	18.4	64	16.6	5,124	18.4
Family history of diabetes						
No	27,878	98.6	373	96.6	27,505	98.6
Yes	401	1.4	13	3.4	388	1.4
Has chronic disease(s)						
No	19,769	70.0	221	57.3	19,548	70.1
Yes	8,510	30.0	165	42.8	8,345	29.9
Smoking status						
Never smoked	21,857	77.3	308	79.8	21,549	77.3
Ever smoked	6,422	22.7	78	20.2	6,344	22.7
Passive smoking exposure						
Not exposed	14,073	49.7	161	41.7	13,912	49.9
Exposed	6,044	21.4	75	19.4	5,969	21.4
Unknown	8,162	28.9	150	38.9	8,012	28.7
Vegetable consumption (kg/week)						
<1.0	2,290	8.1	33	8.6	2,257	8.1
1.0–1.9	12,382	43.8	146	37.8	12,236	43.9
≥2.0	13,109	46.4	198	51.3	12,911	46.3
Unknown	498	1.8	9	2.3	489	1.8
Fruit consumption (kg/week)						
<0.5	9,326	33.0	117	30.3	9,209	33.0
0.5–1.4	13,263	46.9	198	51.3	13,065	46.8
≥1.5	4,828	17.1	58	15.0	4,770	17.1

Baseline characteristics	Total cohort		New cases of T2DM		Participants without T2DM at end of follow-up	
Unknown	862	3.1	13	3.4	849	3.0
Exercise						
No	22,681	80.2	307	79.5	22,374	80.2
Yes	2,526	8.9	31	8.0	2,495	8.9
Unknown	3,072	10.9	48	12.4	3,024	10.8
Body mass index (kg/m^2)						
Normal or underweight (<24.9)	22,378	79.1	226	58.6	22,152	79.4
Overweight (25.0–29.9)	5,530	19.6	134	34.7	5,396	19.4
Obese (≥30)	371	1.3	26	6.7	345	1.2
Education						
Junior high school or higher	5,760	20.4	62	16.1	5,698	20.4
No junior high school diploma	22,229	78.6	322	83.4	21,907	78.5
Unknown	290	1.0	2	0.5	288	1.0
Occupation						
Non-farmer	7,607	26.9	85	22.0	7,522	27.0
Farmer	20,672	73.1	301	78.0	20,371	73.0
Family income						
Low	4,115	14.6	50	13.0	4,065	14.6
Middle	21,954	77.6	300	77.7	21,654	77.6
High	2,210	7.8	36	9.3	2,174	7.8

Abbreviations: T2DM, type 2 diabetes mellitus.
Definitions: Chronic disease(s) included are hypertension, asthma, hemorrhagic stroke, ischemic stroke, coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD), and tumour. Education from elementary school to junior high school (Grade 9 equivalent in Canada) is compulsory in China. Occupations other than farming include public officers, corporate, private employees, business, and others.

Compared to participants who did not develop T2DM at the end of follow-up, the proportion of female new cases of T2DM was higher (62.7% vs. 55.7%), the mean age was 57 years old (SD: 14.7 years), the proportion of those who had a pre-existing chronic disease was higher (42.8% vs. 29.9%), and the mean BMI was 22.5 kg/m^2 (SD: 3.0 kg/m^2). The majority of other baseline characteristic categories did not differ drastically between those who did and did not develop T2DM at the end of follow-up.

The total 28,279 participants contributed to 118,117.3 person-years of follow-up (**Table 3**). The individual person-years of follow-up for each participant did not appear normally distributed, therefore the median is described; the median follow-up was 3.4 years (IQR: 2.1). The incidence of T2DM was 3.27

per 1,000 person-years (95% CI: 2.96, 3.61). The incidence rate is comparable to a study done in Taiwan,[34] but lower than the incidence rate in a nationwide China project.[54] The incidence rate of T2DM was higher in females, with increasing age at baseline, having a family history of T2DM, having a chronic disease, and being overweight or obese.

Table 3. Incidence Rates of Type 2 Diabetes Mellitus, According to Baseline Characteristics of Study Participants

Baseline characteristics	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI
Overall	386	118,117.3	3.27	2.96, 3.61
Gender				
Female	242	65,331.5	3.70	3.27, 4.20
Male	144	52,785.9	2.73	2.32, 3.21
Age at baseline (years)				
18–39	8	20,667.5	0.39	0.19, 0.77
40–49	63	29,806.7	2.11	1.65, 2.71
50–59	143	32,504.3	4.40	3.73, 5.18
60–69	108	21,427.1	5.04	4.17, 6.09
≥70	64	13,711.8	4.67	3.65, 5.96
Family history of diabetes				
No	373	116,462.2	3.20	2.89, 3.54
Yes	13	1,655.2	7.85	4.56, 13.53
Has chronic disease(s)				
No	221	84,254.0	2.62	2.30, 2.99
Yes	165	33,863.3	4.87	4.18, 5.68
Smoking status				
Never smoked	308	88,459.4	3.48	3.11, 3.89
Ever smoked	78	29,657.9	2.63	2.11, 3.28
Passive smoking exposure				
Not exposed	161	41,076.8	3.92	3.36, 4.57
Exposed	75	24,544.2	3.06	2.44, 3.83
Unknown	150	52,496.3	2.86	2.43, 3.35
Vegetable consumption (kg/week)				
<1.0	33	7,676.4	4.30	3.06, 6.05
1.0–1.9	146	44,232.2	3.30	2.81, 3.88
≥2.0	198	64,306.4	3.08	2.68, 3.54
Unknown	9	1,902.3	4.73	2.46, 9.09
Fruit consumption (kg/week)				
<0.5	117	35,025.2	3.34	2.79, 4.00

Baseline characteristics	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI
0.5–1.4	198	55,111.0	3.59	3.13, 4.13
≥1.5	58	24,711.6	2.35	1.81, 3.04
Unknown	13	3,269.5	3.98	2.31, 6.85
Exercise				
No	307	98,926.6	3.10	2.77, 3.47
Yes	31	9,580.9	3.24	2.28, 4.60
Unknown	48	9,609.8	4.99	3.76, 6.63
Body mass index (kg/m^2)				
Normal or underweight (<24.9)	226	95,307.0	2.37	2.08, 2.70
Overweight (25.0–29.9)	134	21,404.0	6.26	5.29, 7.42
Obese (≥30)	26	1,406.4	18.49	12.59, 27.15
Education				
Junior high school or higher	62	32,794.8	1.89	1.47, 2.42
No junior high school diploma	322	84,261.3	3.82	3.43, 4.26
Unknown	2	1,061.1	1.88	0.47, 7.54
Occupation				
Non-farmer	85	33,671.0	2.52	2.04, 3.12
Farmer	301	84,446.3	3.56	3.18, 3.99
Family income				
Low	50	15,850.9	3.15	2.39, 4.16
Middle	300	93,042.9	3.22	2.88, 3.61
High	36	9,223.5	3.90	2.82, 5.41

Abbreviations: T2DM, type 2 diabetes mellitus; CI, confidence interval.
Definitions: Chronic disease(s) included are hypertension, asthma, hemorrhagic stroke, ischemic stroke, coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD), and tumour. Education from elementary school to junior high school (Grade 9 equivalent in Canada) is compulsory in China. Occupations other than farming include public officers, corporate, private employees, business, and others.

For the main analysis, $PM_{2.5}$ concentrations at baseline were assigned to each participant as the proxy measure for personal exposure. The mean baseline exposure to $PM_{2.5}$ was 51.14 μg/m^3 (SD: 5.27 μg/m^3) with an interquartile range (IQR) of 8.92 μg/m^3, ranging from 39.42 to 82.94 μg/m^3 (**Table 4**). Assignment of exposure as the time-fixed, 3-year average (**Table S2.1, Figure S2.1**) or yearly time-varying (**Table S3.1, Figure S3.1**) concentration of $PM_{2.5}$ did not result in drastically differing results.

Table 4. Baseline Concentration of Ambient Fine Particulate Matter

Baseline concentration (μg/m³)	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
$PM_{2.5}$	51.14	5.27	39.42	47.17	50.83	56.08	82.92	8.92

Abbreviations: $PM_{2.5}$, ambient fine particulate matter.

Baseline characteristics of study participants were additionally presented according to quartile category of baseline $PM_{2.5}$ exposure. The 1st through 4th quartiles of $PM_{2.5}$ were: less than 47.17 μg/m³, 47.17 to 50.82 μg/m³, 50.83 to 56.07 μg/m³, and greater than or equal to 56.08 μg/m³, respectively (**Table 5**). Given the nature of quartiles, the numbers of participants were equal amongst each category of $PM_{2.5}$ quartiles for the total cohort. Slightly more participants developed T2DM at the end of follow-up compared to the total cohort in the 2nd quartile (47.17 to 50.82 μg/m³). There were slightly more participants who developed T2DM at the end of follow-up. Baseline characteristics of study participants were presented according to quartiles of baseline $PM_{2.5}$ exposure to determine which covariates may be correlated with $PM_{2.5}$. Covariates that present as being correlated with both T2DM (**Table 2**) and $PM_{2.5}$ (**Table 6**) require consideration as potential confounders of the association of interest.

Table 5. Quartiles of Baseline Exposure to Ambient Fine Particulate Matter, According to Type 2 Diabetes Mellitus Status at End of Study

Baseline exposure	Total cohort		New cases of T2DM		Participants without T2DM at end of follow-up	
	No.	%	No.	%	No.	%
Overall	28,279	100.0	386	100.0	27,893	100.0
$PM_{2.5}$ (μg/m³)						
1st quartile (<47.17)	7,189	25.4	94	24.4	7,095	25.4
2nd quartile (47.17–50.82)	6,895	24.4	119	30.8	6,776	24.3
3rd quartile (50.83–56.07)	6,837	24.2	84	21.8	6,753	24.2
4th quartile (≥56.08)	7,358	26.0	89	23.1	7,269	26.1

Abbreviations: $PM_{2.5}$, ambient fine particulate matter; T2DM, type 2 diabetes mellitus.

Table 6. Baseline Characteristics of Study Participants, According to Quartiles of Baseline Exposure to Ambient Fine Particulate Matter

Baseline characteristics	Total cohort		Quartiles of baseline $PM_{2.5}$ exposure (μg/m^3)							
			1st (<47.17)		2nd (47.17–50.82)		3rd (50.83–56.07)		4th (≥56.08)	
	No.	%	No.	%	No.	%	No.	%	No.	%
Overall	28,279	100.0	7,189	25.4	6,895	24.4	6,837	24.2	7,358	26.0
Diabetes at follow-up										
No	27,893	100.0	7,095	25.4	6,776	24.3	6,753	24.2	7,269	26.1
Yes	386	100.0	94	24.4	119	30.8	84	21.8	89	23.1
Gender										
Female	15,766	100.0	3,908	24.8	3,910	24.8	3,903	24.8	4,045	25.7
Male	12,513	100.0	3,281	26.2	2,985	23.9	2,934	23.5	3,313	26.5
Age at baseline (years)										
18–39	3,372	100.0	947	28.1	1,170	34.7	589	17.5	666	19.8
40–49	5,628	100.0	1,824	32.4	1,767	31.4	1,081	19.2	956	17.0
50–59	7,225	100.0	2,077	28.8	1,893	26.2	1,663	23.0	1,592	22.0
60–69	6,866	100.0	1,362	19.8	1,177	17.1	1,985	28.9	2,342	34.1
≥70	5,188	100.0	979	18.9	888	17.1	1,519	29.3	1,802	34.7
Family history of diabetes										
No	27,878	100.0	7,004	25.1	6,809	24.4	6,798	24.4	7,267	26.1
Yes	401	100.0	185	46.1	86	21.5	39	9.7	91	22.7
Has chronic disease(s)										
No	19,769	100.0	4,825	24.4	5,064	25.6	5,018	25.4	4,862	24.6
Yes	8,510	100.0	2,364	27.8	1,831	21.5	1,819	21.4	2,496	29.3
Smoking status										
Never smoked	21,857	100.0	5,431	24.9	5,307	24.3	5,513	25.2	5,606	25.7
Ever smoked	6,422	100.0	1,758	27.4	1,588	24.7	1,324	20.6	1,752	27.3
Passive smoking exposure										
Not exposed	14,073	100.0	3,825	27.2	2,289	16.3	3,887	27.6	4,072	28.9
Exposed	6,044	100.0	1,166	19.3	1,683	27.9	1,670	27.6	1,525	25.2
Unknown	8,162	100.0	2,198	26.9	2,923	35.8	1,280	15.7	1,761	21.6

Baseline characteristics	Total cohort		Quartiles of baseline $PM_{2.5}$ exposure (μg/m^3)							
			1st (<47.17)		2nd (47.17–50.82)		3rd (50.83–56.07)		4th (≥56.08)	
	No.	%	No.	%	No.	%	No.	%	No.	%
Vegetable consumption (kg/week)										
<1.0	2,290	100.0	1,224	53.5	316	13.8	368	16.1	382	16.7
1.0–1.9	12,382	100.0	2,241	18.1	2,676	21.6	4,267	34.5	3,198	25.8
≥2.0	13,109	100.0	3,612	27.6	3,736	28.5	2,125	16.2	3,636	27.7
Unknown	498	100.0	112	22.5	167	33.5	77	15.5	142	28.5
Fruit consumption (kg/week)										
<0.5	9,326	100.0	3,018	32.4	1,818	19.5	1,880	20.2	2,610	28.0
0.5–1.4	13,263	100.0	2,733	20.6	3,543	26.7	3,630	27.4	3,357	25.3
≥1.5	4,828	100.0	1,248	25.9	1,287	26.7	1,209	25.0	1,084	22.5
Unknown	862	100.0	190	22.0	247	28.7	118	13.7	307	35.6
Exercise										
No	22,681	100.0	6,257	27.6	5,748	25.3	5,151	22.7	5,525	24.4
Yes	2,526	100.0	304	12.0	637	25.2	1,016	40.2	569	22.5
Unknown	3,072	100.0	628	20.4	510	16.6	670	21.8	1,264	41.2
Body mass index (kg/m^2)										
Normal or underweight (<24.9)	22,378	100.0	5,683	25.4	5,565	24.9	5,432	24.3	5,698	25.5
Overweight (25.0–29.9)	5,530	100.0	1,428	25.8	1,253	22.7	1,326	24.0	1,523	27.5
Obese (≥30)	371	100.0	78	21.0	77	20.8	79	21.3	137	36.9
Education										
Junior high school or higher	5,760	100.0	1,889	32.8	1,770	30.7	1,008	17.5	1,093	19.0
No junior high school diploma	22,229	100.0	5,247	23.6	5,061	22.8	5,743	25.8	6,178	27.8
Unknown	290	100.0	53	18.3	64	22.1	86	29.7	87	30.0
Occupation										
Non-farmer	7,607	100.0	1,274	16.8	2,336	30.7	1,795	23.6	2,202	29.0
Farmer	20,672	100.0	5,915	28.6	4,559	22.1	5,042	24.4	5,156	24.9
Family income										
Low	4,115	100.0	1,180	28.7	550	13.4	1,134	27.6	1,251	30.4
Middle	21,954	100.0	5,616	25.6	5,943	27.1	5,001	22.8	5,394	24.6

Baseline characteristics	Total cohort		Quartiles of baseline $PM_{2.5}$ exposure ($\mu g/m^3$)							
			1st (<47.17)		2nd (47.17–50.82)		3rd (50.83–56.07)		4th (≥56.08)	
	No.	%	No.	%	No.	%	No.	%	No.	%
High	2,210	100.0	393	17.8	402	18.2	702	31.8	713	32.3

Abbreviations: T2DM, type 2 diabetes mellitus; $PM_{2.5}$, ambient fine particulate matter.

Definitions: Chronic disease(s) included are hypertension, asthma, hemorrhagic stroke, ischemic stroke, coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD), and tumour. Education from elementary school to junior high school (Grade 9 equivalent in Canada) is compulsory in China. Occupations other than farming include public officers, corporate, private employees, business, and others.

The incidence rates of T2DM according to quartile category of baseline $PM_{2.5}$ exposure additionally indicated that development of T2DM was related to exposure to greater concentrations of $PM_{2.5}$ (**Table 7**).

Table 7. Incidence Rates of Type 2 Diabetes Mellitus, According to Quartiles of Baseline Exposure to Ambient Fine Particulate Matter

Baseline exposure	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI
Overall	386	118,117.3	3.27	2.96, 3.61
$PM_{2.5}$				
1st quartile (<47.17)	94	36,679.4	2.56	2.09, 3.14
2nd quartile (47.17–50.82)	119	38,146.4	3.12	2.61, 3.73
3rd quartile (50.83–56.07)	84	23,225.4	3.62	2.92, 4.48
4th quartile (≥56.08)	89	20,066.1	4.44	3.60, 5.46

Abbreviations: T2DM, type 2 diabetes mellitus; CI, confidence interval; $PM_{2.5}$, ambient fine particulate.

Assessment for Multiplicative Effect Modification

Significance of effect modification was assessed on the multiplicative scale using the cross product interaction term of $PM_{2.5}$ with each hypothesized effect modifying covariate, namely gender, age, smoking status, and BMI. A p-value for interaction of less than 0.05 was considered strong statistically significant evidence for effect modification. Borderline statistical significant results, from a p-value less than 0.10, were still assessed in further steps. When $PM_{2.5}$ exposure assigned at baseline was implemented in the model, there was only borderline statistical significant evidence of effect modification for smoking status and BMI, with p-values of 0.0688 and 0.0708, respectively (**Table 8.1**). When exposure was assigned as time-fixed, 3-year average (**Table S2.2**) concentration of $PM_{2.5}$, multiplicative interaction terms with smoking status and BMI similarly exhibited the greater significance, specifically with p-values of 0.0136 and 0.0045, respectively. When exposure was assigned as yearly time-varying (**Table S3.2**) concentration of $PM_{2.5}$, only the cross product of $PM_{2.5}$ with smoking status was statistically significant, with a p-value of 0.0039.

Table 8.1. Effect Modification of Baseline Exposure to Ambient Fine Particulate Matter and Selected Baseline Characteristics on Incidence of Type 2 Diabetes Mellitus

Model	Variables in baseline exposure model	β	SE	p-value for interaction
1	**Unadjusted model**: $PM_{2.5}$	—	—	—
2	$PM_{2.5}$ + Gender + ($PM_{2.5}$ × gender)			0.7292
	Male vs. Female	0.0068	0.0195	
3	$PM_{2.5}$ + Age + ($PM_{2.5}$ × Age)			0.1793
	<40 vs. 50–59	-0.2128	0.1021	
	40–49 vs. 50–59	-0.0016	0.0291	
	60–69 vs. 50–59	0.0259	0.0250	
	≥70 vs. 50–59	0.0202	0.0308	
4	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0688
	Ever smoked vs. Never smoked	0.0390	0.0215	
5	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)			0.0708
	Overweight vs. Normal or underweight	-0.04281	0.02048	
	Obese vs. Normal or underweight	-0.05260	0.03911	

A p-value for interaction of <0.05 was considered statistically significant and an indicator of potential effect modification. Reference categories were chosen based on category with largest sample size in order to ensure stable results. All covariates presented as potential effect modifiers were selected a priori based on the current literature. Although smoking status and BMI did not make the cut-off for strong evidence of effect modification, given that they were identified as effect modifiers in the other exposure models, they were assessed and presented in Table 8 and Table 10.
*Abbreviations: β, regression coefficient for $PM_{2.5}$; SE, standard error; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (*kg/m^2*), categorized as normal or underweight (<24.9), overweight (25.0–29.9), and obese (≥30).*

Model fit for the inclusion of two effect modifiers over one was tested using the likelihood ratio test, for which a p-value of less than 0.05 was considered statistically significant. Based on the results of the likelihood ratio test, the inclusion of interaction terms for both BMI with $PM_{2.5}$ and smoking status with $PM_{2.5}$ did not significantly improve the model fitting (**Table 8.2**), indicated by a p-value of 0.06670. However, when exposure was assigned as the 3-year average concentration of $PM_{2.5}$, the inclusion of two effect modifiers (i.e. both BMI with $PM_{2.5}$ and smoking status with $PM_{2.5}$) demonstrated improvement to the model, with a p-value of 0.0260 (**Table S2.3**). As a result of the findings of the likelihood ratio test, models using exposure assigned as baseline $PM_{2.5}$ concentration should be presented stratified by smoking status and stratified by both smoking status and BMI for models using exposure assigned as 3-year average $PM_{2.5}$. Given only one interaction term was statistically significant

for exposure assigned as time-varying $PM_{2.5}$ (**Table S3.2**), no likelihood ratio test was required, and models only need to be stratified by smoking status.

Table 8.2. Likelihood Ratio Tests for the Inclusion of Two Effect Modifiers of the Association Between Incidence of Type 2 Diabetes Mellitus and Baseline Exposure to Ambient Fine Particulate Matter

Model	Variables in baseline exposure model	-2 LOG L	*df*	LR test	p-value
1	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)	7,419.024	5	—	—
2	$PM_{2.5}$ + BMI + smoking status + ($PM_{2.5}$ × BMI) + ($PM_{2.5}$ × smoking status)	7,413.609	7	5.415	0.06670

*A p-value for the likelihood ratio test of <0.05 was considered statistically significant and an indicator of better model fit. Abbreviations: -2 LOG L, -2 log-likelihood (deviance); df, degrees of freedom; LR test, likelihood ratio chi–square test; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (*kg/m^2*), categorized as normal or underweight (<24.9), overweight (25.0–29.9), and obese (≥30).*

Adjustment for Potential Confounders

Models were confounder-adjusted at three degrees of adjustment: unadjusted, parsimoniously-adjusted, and fully-adjusted. The unadjusted model includes only $PM_{2.5}$ exposure as the independent variable, and T2DM incidence as the dependent variable. The parsimoniously-adjusted model includes all covariates demonstrating statistical evidence of confounding. The parsimoniously-adjusted model was prepared in consideration for the potential issue of overfitting.[41,42] Percent change of regression coefficient was calculated as |(nested model β – fuller model β) ÷ fuller model β|×100%. A percent change of regression coefficient of greater than or equal to 10% was considered an "important change"[26] and evidence of confounding of the association. All covariates presented as potential confounders were selected *a priori* based on the current literature. The parsimoniously-adjusted model included all covariates that had a percent change of regression coefficient greater than or equal to 10%. The fully-adjusted model includes all covariates identified *a priori* as potential confounders.

The percent change criteria for inclusion into the parsimoniously-adjusted model are presented in **Table 9**. Inclusion of age, BMI, and education all led to a percent change in β of $PM_{2.5}$ greater than

10%, suggesting statistical evidence of confounding, and are therefore included in the parsimoniously-adjusted model. Tables of calculated percent changes in β of $PM_{2.5}$ for 3-year average and time-varying exposure are presented in **Table S2.4** and **Table S3.3**, respectively. For the 3-year average exposure model, age, passive smoking status, vegetable consumption, history of disease, and occupation exhibited evidence of confounding, and are included in the parsimoniously-adjusted 3-year average model. For the time-varying exposure model, age, education, and occupation exhibited evidence of confounding, and are included in the parsimoniously-adjusted time-varying model.

Table 9. Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Baseline Exposure to Ambient Fine Particulate Matter, Adjusted Independently By Potential Confounders

Model	Variables in baseline exposure model	HR	95% CI	p-value	β	Percent change
1	Unadjusted model: $PM_{2.5}$	1.37	1.13, 1.67	0.0014	0.03156	—
2	$PM_{2.5}$ + Gender	1.37	1.13, 1.67	0.0015		0.16
3	$PM_{2.5}$ + Age	1.26	1.03, 1.54	0.0232		36.21
4	$PM_{2.5}$ + BMI	1.33	1.09, 1.61	0.0042		11.21
5	$PM_{2.5}$ + Exercise	1.35	1.11, 1.65	0.0024		4.37
6	$PM_{2.5}$ + Smoking status	1.37	1.13, 1.66	0.0016		0.67
7	$PM_{2.5}$ + Passive smoking exposure	1.35	1.11, 1.64	0.0027		5.09
8	$PM_{2.5}$ + Vegetable consumption	1.41	1.16, 1.71	0.0006		7.67
9	$PM_{2.5}$ + Fruit consumption	1.35	1.11, 1.64	0.0025		4.64
10	$PM_{2.5}$ + Family history of diabetes	1.39	1.15, 1.69	0.0008		4.42
11	$PM_{2.5}$ + History of disease	1.37	1.13, 1.66	0.0013		0.22
12	$PM_{2.5}$ + Education	1.32	1.09, 1.61	0.0053		12.71
13	$PM_{2.5}$ + Occupation	1.41	1.16, 1.70	0.0005		7.23
14	$PM_{2.5}$ + Family income	1.37	1.13, 1.66	0.0017		1.09
15	**Parsimoniously-adjusted model**[a]	1.22	1.00, 1.49	0.0523	—	—

Percent change of regression *coefficient was calculated as |(nested model* β *– fuller model* β*) ÷ fuller model* β*|×100%. A percent change of* regression *coefficient ≥10% was considered strong evidence of confounding of the association between long-term exposure to ambient fine particulate matter and incidence of type 2 diabetes mellitus. All covariates presented as potential confounders were selected a priori based on the current literature. The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%.*

[a]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, BMI, and education.

Abbreviations: HR, hazard ratio, per 10-μg/m³ increase in $PM_{2.5}$; CI, confidence interval; β*, regression coefficient; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index.*

Effect Modification in Confounder-Adjusted Models

Evidence of effect modification identified in unadjusted models in a previous step, were then retested in parsimoniously- and fully-adjusted models, to determine whether those covariates remained effect modifiers of the main association of interest. These results are presented in **Table 10**. Specifically, the unadjusted baseline $PM_{2.5}$ exposure model identified smoking status as an effect modifier of the association between $PM_{2.5}$ exposure and incidence of T2DM, the unadjusted 3-year average $PM_{2.5}$ exposure model identified smoking status and BMI as effect modifiers, and the unadjusted time-varying $PM_{2.5}$ exposure model identified smoking status as an effect modifier. When adjusted parsimoniously for age, BMI, and education in the baseline $PM_{2.5}$ exposure model, smoking status only presented borderline evidence of effect modification (p-value for an interaction = 0.0709). When fully-adjusted for all potential confounders, namely gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income, smoking status remained borderline significant (p-value for an interaction = 0.0688).

Table 10. Effect Modification of Baseline Exposure to Ambient Fine Particulate Matter in Parsimoniously- and Fully-Adjusted Models

Model	Variables in baseline exposure model	β	SE	p-value for interaction
1	**Parsimoniously-adjusted model**[a]	—	—	—
2	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0709
	Ever smoked vs. Never smoked	0.03967	0.02197	
3	**Fully-adjusted model**[b]			
4	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0688
	Ever smoked vs. Never smoked	0.03967	0.02180	

[a]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, BMI, and education.
[b]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

The same retesting was performed for models using 3-year average $PM_{2.5}$ exposure (**Table S2.5**) and yearly time-varying $PM_{2.5}$ exposure (**Table S3.4**). When adjusted parsimoniously for age, passive smoking status, vegetable consumption, history of disease, and occupation in the 3-year average $PM_{2.5}$

exposure model, smoking status and obesity remained statistically significant at a p-value for an interaction of 0.0242 and 0.0038, respectively. Full adjustment did not noticeably alter significance. When parsimoniously-adjusted for age, education, and occupation in the time-varying $PM_{2.5}$ exposure model, smoking status remained significant (p-value for an interaction = 0.0057). Similarly, full adjustment did not alter significance in the time-varying exposure model.

Final Adjusted and Stratified Models

The results of the final models are presented in **Table 11** and visually in **Figure 4**. In the unadjusted model, long-term exposure to $PM_{2.5}$ was positively associated with increased risk of developing T2DM and was statistically significant. When the model was parsimoniously- or fully-adjusted, the association between long-term $PM_{2.5}$ exposure and development of T2DM was slightly attenuated, but remained statistically significant. For each 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure, the HR of T2DM incidence was 1.37 (95% CI: 1.13, 1.67), without adjustment for potential confounding covariates (i.e. unadjusted model). The HR of T2DM incidence for a 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure was 1.22 (95% CI: 1.00, 1.49) when adjusting for only covariates that demonstrated statistical evidence of confounding (i.e. parsimoniously-adjusted model), namely age, BMI, and education. The HR of T2DM incidence for a 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure was 1.26 (95% CI: 1.02, 1.54) when adjusting for all important potential confounding covariates chosen *a priori* (i.e. fully-adjusted model), namely gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Table 11. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Baseline Exposure to Ambient Fine Particulate Matter, Stratified by Smoking Status

Variables in baseline exposure model	Overall		Smoking Status			
			Non-smokers		Smokers	
	(No. of cases = 386)		(No. of cases = 308)		(No. of cases = 78)	
	HR	95% CI	HR	95% CI	HR	95% CI
Unadjusted model[a]	1.37	1.13, 1.67	1.24	0.99, 1.55	1.84	1.27, 2.66
Parsimoniously-adjusted model[b]	1.22	1.00, 1.49	1.11	0.88, 1.40	1.65	1.16, 2.45
Fully-adjusted model[c]	1.26	1.02, 1.54	1.16	0.91, 1.47	1.68	1.12, 2.52

The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%. The fully-adjusted model includes all covariates chosen a priori based on the current literature.

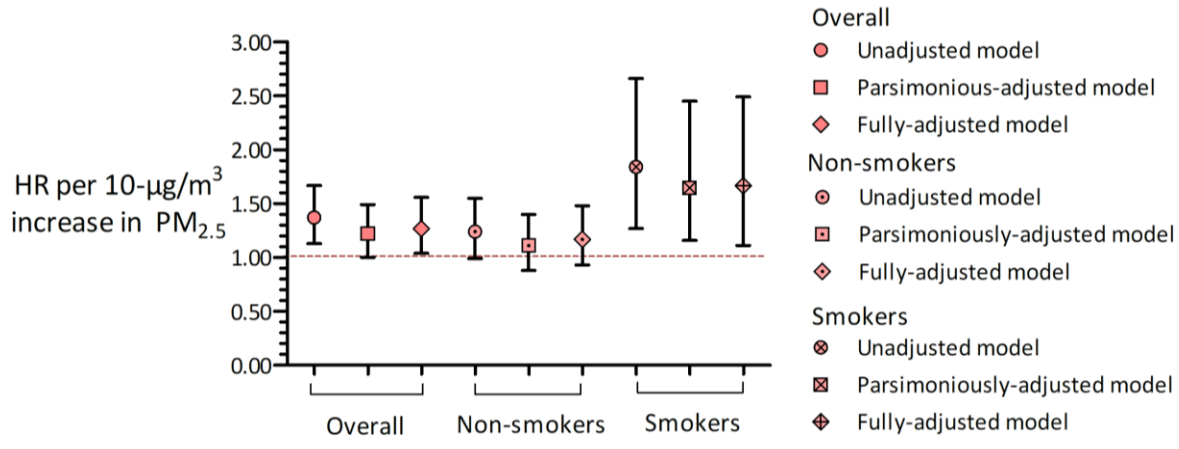
[a]Covariates in the unadjusted model include $PM_{2.5}$ only.

[b]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, BMI, and education.

[c]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (kg/m^2).

Figure 4. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Baseline Exposure to Ambient Fine Particulate Matter, Stratified by Smoking Status



Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$ with 95% confidence intervals.

When stratified by smoking status, the effect modifier of the association of interest, the risk of developing T2DM increased for participants who were ever smokers compared to non-smokers. The association was similarly higher in the unadjusted model for smokers, and slightly attenuated when parsimoniously- or fully-adjusted, but maintained statistical significance. For instance, in the fully-

adjusted model, each 10-μg/m^3 increase in $PM_{2.5}$ exposure was associated with a 68% increased risk for T2DM incidence in smokers (HR = 1.68, 95% CI: 1.12, 2.52). In comparison, $PM_{2.5}$ exposure was associated with a 16% increased risk in non-smokers (HR = 1.16, 95% CI: 0.91, 1.47). However, the association in non-smokers was not statistically significant (contained the null HR = 1.00), suggesting that no evidence of the association in non-smokers and therefore inferences regarding non-smokers cannot be made with confidence.

Similar results were calculated when using 3-year average or yearly time-varying $PM_{2.5}$ exposure. When exposure was assigned as the time-fixed, 3-year average $PM_{2.5}$, T2DM incidence remained positively associated with $PM_{2.5}$ exposure (**Tables S2.4–S2.7, Figures S2.2–S2.3**). In the fully-adjusted model, each 10-μg/m^3 increase in $PM_{2.5}$ exposure was associated with a 43% increased risk for T2DM incidence (HR = 1.43, 95% CI: 1.07, 1.89). Recall that smoking status and BMI both exhibited traits of effect modification, and are therefore hereon presented in stratified analysis. When stratified by smoking status, smokers exhibited increased risk of developing T2DM associated with $PM_{2.5}$ exposure, just as was found when using baseline exposure to $PM_{2.5}$ (**Table S2.6, Figure S2.2**); in the fully-adjusted model, the HR for T2DM incidence for a 10-μg/m^3 increase in $PM_{2.5}$ exposure was 2.66 (95% CI: 1.48, 4.81). Comparatively, the association in non-smokers was not statistically significant, and so inferences cannot be made for the positive association between $PM_{2.5}$ and incident T2DM. When stratified by BMI, only the category of normal or underweight presented positive associations with statistical significance (**Table S2.7, Figure S2.3**); the categories of overweight and obese both were not statistically significant. The association between $PM_{2.5}$ exposure and development of T2DM tended to be negative. However, given the lack of statistical significance, such an interpretation remains inconclusive.

When exposure was assigned as the yearly time-varying $PM_{2.5}$, T2DM incidence remained positively associated with $PM_{2.5}$ exposure. In the fully-adjusted model, each 10-μg/m^3 increase in $PM_{2.5}$

exposure was associated with a 51% increased risk for T2DM incidence (HR = 1.51, 95% CI: 1.14, 1.99). When stratified by smoking status, smokers exhibited increased risk of developing T2DM associated with $PM_{2.5}$ exposure (**Table S3.5, Figure S3.2**); in the fully-adjusted model, the HR for T2DM incidence associated with a 10-μg/m^3 increase in $PM_{2.5}$ exposure was 2.42 (95% CI: 1.36, 4.30). Similar to the model using baseline exposure to $PM_{2.5}$, in this model, the association in non-smokers lost statistical significance, suggesting lack of strong enough evidence to make a confident inference. The relationship between the effect modification by smoking status on the association between $PM_{2.5}$ and T2DM exhibited similar results regardless of the exposure metric, baseline, 3-year average, or time-varying.

Sensitivity Analyses

Analyses using different exposure metric assignments of $PM_{2.5}$ suggested that the main results were sensitive to assignment choice. The main analysis used $PM_{2.5}$ exposure assigned at baseline as a proxy measure for "true" personal long-term exposure to $PM_{2.5}$. Other possible exposure metrics assigned in other studies of long-term exposure include: multi-year average, yearly time-varying, multi-year moving average, full period average, yearly time-varying with lag, and multi-year moving average with lag. The results of this sensitivity analysis suggest that baseline assigned $PM_{2.5}$ results in HRs similar in magnitude and precision as its multi-year average or time-varying counterparts (**Table S4.1, Figure S4.1**). Compared to the models using time-varying $PM_{2.5}$ exposures, the time-fixed, baseline exposure resulted in slightly attenuated associations. For example, the unadjusted HR for exposure assigned at baseline was 1.37 (95% CI: 1.13, 1.67), compared to 1.55 (95% CI: 1.19, 2.02) for exposure assigned as a yearly time-varying variable, and 1.53 (95% CI: 1.15, 2.05) for exposure assigned as a 3-year moving average. Compared to models using multi-year average exposures, baseline exposure led to slightly tightened CI, but exhibited no pattern of estimating over or under the multi-year average exposures. For

time-varying exposures, assigned a single-year or multi-year moving average, additionally did not affect the results drastically; the HR for exposure assigned as a yearly time-varying variable was 1.55 (95% CI: 1.19, 2.02) and 1.53 (95% CI: 1.15, 2.05) for exposure assigned as a 3-year moving average. Compared to models implementing a 1-year lag, non-lagged exposures did not drastically alter the findings. Overall, the sensitivity analyses on differing assignment of $PM_{2.5}$ exposure suggest that baseline exposure, although may lead to slightly attenuated results compared to multi-year or time-varying assignments, approximates true exposure to an sufficient degree.

The second form of sensitivity analyses was performed with regards to inclusion of participants into the cohort. The original analytical cohort includes all participants from the original Rural Deqing Cohort, excluding only those with T2DM at baseline (prevalent cases), or missing data on key covariates, namely $PM_{2.5}$ exposure, gender, and smoking status. In the first subset cohort for the sensitivity analysis, all participants that developed T2DM in the first year of follow were excluded, resulting in the reduction to 309 cases identified. The purpose of this first subset cohort is to remove participants with underlying risk of T2DM independent from $PM_{2.5}$ exposure levels.[54,59] In the second subset cohort for the sensitivity analysis, participants who contributed less than or equal to two person-years of follow-up were excluded, resulting in a reduction to 167 cases identified.

By excluding cases developed in their first year of follow-up, the risk of T2DM incidence for a 10-µg/m3 increase in $PM_{2.5}$ using the parsimoniously-adjusted model was positive and statistically significant (HR = 1.27; 95% CI: 1.02, 1.59) and therefore comparable to the original main analysis (**Table S4.2, Figure S4.2**). Similar risks were found when exposure was assigned as a 3-year average and a yearly time-varying variable. When participants contributing to two years of follow-up or less were excluded, risk of T2DM incidence was more strongly associated with $PM_{2.5}$ exposure (HR = 1.60; 95% CI: 1.20, 2.14).

One final sensitivity analysis was performed for the definition of obesity used in the supplementary analysis for models using exposure assigned as a 3-year average. Obesity, defined by categories of BMI, was determined to be an effect modifier of the relationship between $PM_{2.5}$ exposure assigned as a 3-year average, and T2DM incidence. The BMI cut-offs for normal, overweight, and obese, were chosen based on WHO guidelines. In this sensitivity analysis, BMI was recategorized based on Chinese guidelines (**Table S4.3**). Asians are known to have lower rates of overweight and obesity compared to Western people when using WHO definitions; however, research confirms that for Asians, the risk of T2DM starts at a lower BMI than for Western people.[44,45,69] For the Chinese standard, normal weight is considered a BMI less than 23, compared to 25 for WHO.[23] Overweight is a BMI anywhere between 23 and 27.5, and obese is defined as a BMI of 27.5 or greater, a much lower cut-off compared to obese defined by WHO (i.e. BMI of 30 or higher). The distribution of participants by BMI category was different when using Chinese standards; more participants were considered overweight (35.32%) compared to when overweight was defined using WHO standards (19.6%). When using the Chinese standards for BMI cut-offs for normal, overweight, and obese, the association in normal or underweight individuals remains positive and statistically significant, overweight became positive but remained insignificant, and obese was attenuated closer to the null but remained negatively associated and statistically insignificant (**Table S4.4, Figure S4.4**–**S4.5**). Additionally, in a subset cohort of participants with data on measurements of abdominal fat, alternative dichotomous covariate for obesity was implemented, defined by cut-offs for obesity based on WHO and Chinese guidelines on abdominal fat (**Table S4.5**). Research has shown that waist circumference (cm), a measure of abdominal fat, is a better measure for obesity in Chinese people as a risk factor for T2DM.[45,70] Waist circumference has previously been utilized as a measure of obesity in previous studies on prevalence of T2DM in China.[66] This subset cohort was reduced to 11,457 total participants, and 193 identified new cases of T2DM. When using waist circumference as a measure of obesity, similar to the main analysis using WHO standards for

categorized BMI as a measure of obesity, it demonstrated positive associations for non-obese individuals, and negative associations for obese individuals, but without statistical significance (**Table S4.6, Figure S4.5**). Altogether, the sensitivity analysis for obesity, specifically using Chinese standards for categorized BMI as a measure of overall obesity, and using WHO and Chinese standards for categorizing waist circumference as a measure of abdominal obesity, indicated that the findings of the main analysis, using WHO standards for categorized BMI, cannot hold any strong conclusions, as the differences in associations by degree of obesity changed based on the choice of proxy for personal obesity.

Model Assumptions

A fundamental assumption for using the Cox proportional hazards model is the proportional hazards.[41] Whether the proportional hazards assumption was true for exposure assigned as baseline concentrations of $PM_{2.5}$ was assessed. A log of negative log survival plot of quartiles of $PM_{2.5}$ against time was developed to visually assess for evidence of nonproportionality (**Figure S5.1**). Patterns to look for in this plot are convergence, divergence, or crossing of curves, for which converging curves suggest decreased difference over time, diverging curves suggest increased difference over time, and pronounced crossing of curves suggest the nonproportionality is important to be considered when making final inferences.[41] For the most part, the curves of $PM_{2.5}$ quartiles remained tightly closer together, over the beginning years of the follow-up period suggesting no issue of nonproportionality; however by around year three, curves representing the 1st quartile begin demonstrating crossover with other quartile curves, suggesting nonproportionality may be an issue in the later years of the follow-up period. This violation of the proportional hazards assumption is further indicated by the results of the log of negative log survival plot of quartiles of $PM_{2.5}$ against log of time. Additionally, time was included

as an interaction term into the model as another method of assessing the assumption of proportional hazards, for which a statistically significant coefficient of the product between time and $PM_{2.5}$ suggesting nonproportionality.[71,72] The resulting p-value was significant (p-value = 0.0009), suggesting evidence that the proportional hazards assumption is violated, as $PM_{2.5}$ varies with time (**Table S5**). Finally, plots of the cumulative Martingale residuals against the values of $PM_{2.5}$, via the ASSESS statement in SAS, can additionally be used to assess the proportional hazards assumption, for which a significant p-value suggests violation of the proportionality assumption.[40,73] The p-value was not significant (p-value = 0.1740) suggesting that the proportional hazards assumption was not violated (**Figure S5.2**). Altogether, these findings suggest that the time-varying models of $PM_{2.5}$ exposure may better fit the assumptions needed to be made to run a Cox proportional hazards model. One of the benefits of modelling exposure as a time-varying covariate, is the relaxation model fit assumptions and the ability to model nonproportionality.[40] The time-varying model no longer varies with time in the same proportions when only baseline, time-fixed covariates are used.[26] This has led some authors to call the model a time-varying Cox hazards model, rather than a time-varying Cox proportional hazards model; in other words, the hazard of T2DM incidence at time *t* is no longer proportional to the baseline hazard.[72]

3.5 Discussion

Summary

In this cohort study made up of adult residents of rural Deqing County, Zhejiang, China, long-term exposure to $PM_{2.5}$ was associated with increased risk for incident T2DM, with adjustment for potential confounders. Specifically, the HR for incidence of T2DM associated with per 10-$\mu g/m^3$ in $PM_{2.5}$ increase was 1.22 (95% CI: 1.00, 1.49), when adjusted parsimoniously for age, BMI, and education, and the HR was 1.26 (95% CI: 1.02, 1.54) in the fully-adjusted model. There was evidence of some effect

modification by selected covariates, namely smoking status for the main analysis, in which smokers were at greater risk of T2DM associated with $PM_{2.5}$ exposure. The positive association was consistent in sensitivity analyses by differing temporal metrics of exposure assignment in the statistical model, as well as analyses by redefining cohorts.

Comparison to Other Research

The findings of this study provide epidemiologic evidence of a positive association between long-term exposure to $PM_{2.5}$ and incidence of T2DM. The association between long-term exposure to $PM_{2.5}$ and risk of T2DM has been explored in several cohort studies from North America and Europe, with very few done in Asia.

Of the four cohort studies done in Asian countries, all found positive associations with statistical significance.[25,34,54,55] In a study including four large-scale cohorts in China, followed up for a total of 11 years, the adjusted HR for incidence of T2DM per 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure was 1.16 (95% CI: 1.06, 1.26).[54] Comparably, two separate cohort studies from Taiwan also found positive associations, with an adjusted HR for incidence of T2DM comparing the 4th quartile of $PM_{2.5}$ exposure to the first of 1.16 (95% CI: 1.07, 1.26), and an adjusted HR for incidence of T2DM per 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure of 1.11 (95% CI: 1.08, 1.13).[34,55] Another cohort study conducted in Hong Kong found a modestly stronger association, with an adjusted HR for incidence of T2DM per 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure of 1.55 (95% CI: 1.18, 2.03).[25]

Furthermore, systematic reviews and meta-analyses (SR-MA) on the association between long-term $PM_{2.5}$ exposure and T2DM reported positive pooled effect estimates. There were three SR-MAs published most recently, during the years of 2019 and 2020.[74–76] One SR-MA published in 2019

presented pooled HR for T2DM incidence per 10-μg/m^3 increase in $PM_{2.5}$ of 1.10 (95% CI: 1.04, 1.16) with substantial heterogeneity (I^2 = 64%, p-value = 0.0009).[74] Another SR-MA published in 2019 showed that the pooled HR per 10-μg/m^3 increase in $PM_{2.5}$ was 1.11 (95% CI: 1.03, 1.19) with substantial heterogeneity (I^2 = 55.8%, p-value = 0.012).[75] Most recently, a SR-MA published in 2020 found that the pooled HR for T2DM incidence per 10-μg/m^3 increase in $PM_{2.5}$ of 1.10 (95% CI: 1.04, 1.17) with substantial heterogeneity (I^2 = 74.4%, p-value = 0.000).[76] None of these SR-MAs were able to meta-analyse the HR for incidence of T2DM specific to Asian countries.

A number of studies assessed the potential for effect modification by various covariates, such as gender, age, BMI, smoking status, and education. Evidence of effect modification in these studies was variable. A cohort study of residents of Ontario, Canada assessed the potential for effect modification by age, sex, and income, and found no evidence for the effect modification.[49] A study of adults aged 50 to 71 years old in the USA assessed effect modification by age, gender, race, smoking history, pre-diagnosed health status, BMI, and diet, finding evidence of positive effect modification by greater BMI and lower fruit consumption.[77] Another study from the US assessed effect modification by gender, ethnicity, education, and obesity with no evidence of an effect.[50] In Europe, all four cohort studies assessed potential for effect modification.[51–53,78] Two cohort studies, one in female nurses in the Denmark,[51] and the other in residences of the Ruhr area of Germany,[53] found positive effect modification by non-smoking status, obesity, and CVD, and male sex and obesity, respectively. In Asia, all four cohort studies assessed potential effect modification of the relationship.[25,34,54,55] The study consisting of four large-scale cohorts from China demonstrated evidence of positive effect modification by being younger, female, living in rural regions, lower BMI, non-smoking status, and non-hypertensive.[54] Interestingly, the present study presented evidence that smoking was a positive effect modifier, unlike the nationwide study. Of the two studies from Taiwan, one prospective cohort study found evidence of positive effect modification by lower BMI and alcoholism,[34] and the other

retrospective cohort study found evidence of positive effect modification by being male, older age, and hyperlipidemia.[55] The study of elderly Hong Kong residents found positive effect modification by being female.[25]

To our knowledge, this is the first cohort study that assessed the association between long-term exposure to $PM_{2.5}$ and incidence of T2DM in a rural Chinese adults. In general, the findings of this cohort study are consistent with the epidemiologic evidence found in other cohorts studying the association between long-term exposure to $PM_{2.5}$ and incidence of T2DM. Potential inconsistencies amongst the differing cohort studies for this association and its effect modifiers may be due to differences in study regions' $PM_{2.5}$ concentrations, mixtures, and sources as well as differences in study populations' underlying risks and baseline population characteristics; for instance, the present rural cohort exhibited very low smoking rates in females and relatively low prevalence of obesity and family history of diabetes overall.

Biologic Plausibility

There have been various hypotheses regarding potential biologic mechanisms underlying the effect of $PM_{2.5}$ exposure on development of T2DM. $PM_{2.5}$ deposits deep within the alveoli of the lungs and can interact with and modify cellular- and molecular-level structures.[11] Some major mechanisms by which $PM_{2.5}$ may induce development of T2DM include endothelial dysfunction, oxidative stress and systemic inflammation, hepatic insulin resistance and stress, and mitochondrial and adipose tissue dysfunction,[2,14,79–82] which have been described in detail in the main body of the manuscript.

Strengths

There were many strengths of the current study that build confidence in the final results. Firstly, covariates included in the fully-adjusted model were chosen *a priori* based on the current literature on T2DM development. Important individual factors, such as BMI, physical activity, and family history of diabetes, were collected and adjusted for as potential confounders of the association.[44–46,83] BMI was calculated using anthropomorphic measurements of height and weight by a physician, preventing the possibility of social desirability bias which may arise in self-reports.

Exposure was assessed at a fine spatial scale of 1×1 km resolution, allowing for more precise exposure assignments to be made compared to coarser resolutions (10×10 km), which may be more suitable for covering larger, national cohort studies.[84]

One major strength of the present study is the analysis of results using differing exposure metrics assigned into the analytical model, namely baseline exposure, 3-year average exposure, yearly time-varying exposure, and the various additional metrics commonly used in epidemiologic research, presented in the sensitivity analysis. Baseline exposure is sometimes the more appropriate model for evaluating chronic disease in comparison to averaged exposure. For example, Chen et al. (2016) realized that, due to PM_{10} concentration decreasing steadily over time, the effect estimate for lung cancer in this region was overestimated when using an averaged concentration, as survivors were assigned lower exposure values as opposed to participants who had died during the study.[64] Shorter timeframes of $PM_{2.5}$ measures have been reported to adequately represent longer-term exposure.[31] On the other hand, average exposure also have its fair share of benefits. For instance, some researchers choose to use averaged exposures over single-year exposures as a proxy measure for true exposure in order to reduce the noise in the annual satellite-derived values.[21] Research on the temporal and spatial metrics for $PM_{2.5}$ associated with mortality suggested that longer moving averages results in stronger associations.[84]

Furthermore, some simulation research suggested that metrics using longer averages were more robust and suffered less bias.[85]

The decision to examine data using time-fixed or time-varying exposure is another factor to consider. One study on $PM_{2.5}$ associated with mortality demonstrated that time-fixed analyses were comparable to their time-varying counterparts.[86] In another simulation study, researchers noted that in general, the moving average metric tended to overestimate the true HR, whereas the time-fixed, 3-year average and baseline $PM_{2.5}$ exposure, underestimated the true hazard ratio.[85] Despite the lack of unified opinion on the best suitable temporal metric of $PM_{2.5}$ exposure, the sensitivity analysis, though we demonstrated that regardless of metric, long-term $PM_{2.5}$ exposure was associated with T2DM incidence, suggests that although we cannot confidently conclude the magnitude of the association, the association does exist.

The majority of studies in China on long-term exposure to $PM_{2.5}$ and T2DM are cross-sectional, therefore only having data on T2DM prevalence, and not incidence.[74–76] Of the cohort studies, all pertain to urban regions of China. This study based on a unique cohort presents a new perspective to the relationship between $PM_{2.5}$ and T2DM, suggesting that the relationship permeates the urban-rural divide, additionally affecting residents of rural regions of China. Furthermore, the use of satellite-derived $PM_{2.5}$ data allowed the study to overcome limitations of assessing the relationship in regions without monitoring stations.[9,63,87] This is important particularly for rural regions of China, which are less likely to have monitoring stations in place; the Chinese National Real-Time Air Quality Network is skewed towards urban regions, which often have several sites per city, and none in rural regions.[88]

Limitations

Like the majority of studies examining this relationship, the present study assumed that all incident cases of diabetes were T2DM because T2DM accounts for 90 to 95% of people with diabetes globally.[89] This may slightly overestimate the true incidence rate of T2DM. Furthermore, the year of diagnosis was assumed to be the year of incidence, additionally introducing potential uncertainties in true date of T2DM incidence.

FPG was used to define and diagnose T2DM in the current study. Other diagnostic tools for T2DM include 2-hour plasma glucose in an oral glucose tolerance test, random plasma glucose, and HbA1c tests.[89–91] HbA1c tests are commonly used in other cohort studies on development of T2DM globally, but are not yet endorsed in China.[90] Each test for T2DM comes with its fair share of disadvantages; for example, results from the FPG test are less tightly linked to diabetes complications than results of the HBA1c test.[92] Ideally, diagnosis would be validated in another dataset or with another diagnostic tool, which was not done for the present study.

Other limitations arise with regards to the covariates measured. For instance, nearly all covariates were taken from self-reported questionnaires at baseline, other than a few anthropomorphic measurements. Recall bias is unavoidable, given a questionnaire was used to collect data on the majority of covariates. Self-reporting bias may be present in certain individual factors, such as smoking status, which have socially desirable responses. However, smoking is largely accepted as social norm in China, particularly amongst adult males, suggesting that social desirability bias in self-reports of smoking status in China are likely minimal.[93]

Additionally, the accuracy of some of the covariate measurements is debatable, particularly given the unique characteristics of rural cohorts. For example, family history of T2DM was extremely low, which may either be reflective of true prevalence, or more likely, be reflective of unknowns in

family history. This is particularly true for the rural Chinese population, as it has been shown that the awareness rates of T2DM are much lower among rural Chinese compared to urban counterparts.[94] In addition to underestimated prevalence of family history, this may lead to more participants testing positive for T2DM in the first years of the study period unrelated to exposure to $PM_{2.5}$. However, such an issue was considered by assessing sensitivity of results by re-examining the results using a subcohort removing diabetes cases identified in the first two years of follow-up.

Another important covariate that may not represent true individual factors of interest is BMI. In the present study, BMI was assumed to be an adequate proxy measure to adiposity. However, research has shown that, particularly, for Asian populations, BMI cut-offs for increased risk of T2DM are lower.[44,45,69,95] Moreover, it has been suggested that BMI itself may not be a suitable measure of adiposity for this population. Waist circumference was also used to measure central obesity, which is considered to be a better measure,[45,70] in a sensitivity analysis of the present study.

Some major sources of $PM_{2.5}$ include the combustion of coal, oil, gasoline, diesel fuel, and wood, atmospheric transformation of products of nitrogen oxides, SO_2, and organic compounds, and high-temperature processes, smelters and steel mills.[7] A major limitation is that source of $PM_{2.5}$ could not be determined from the satellite-derived data. Despite lack of data necessary to determine source of $PM_{2.5}$, educated assumptions can be made based on common major source contributors of similar regions. For instance, in the Health Effects Institute (HEI) Special Report 20, "Burden of Disease Attributable to Coal-Burning and Other Air Pollution Sources in China", the working group found that coal-burning by industry, power plants, and residential heating, were the most important contributors to ambient $PM_{2.5}$ in China. According to the report, in Zhejiang, the province in which Deqing County resides, the major sources of ambient $PM_{2.5}$ included industrial sources from coal emissions, biomass combustion, transport, industrial sources of non-coal emissions, and power plants.[96]

Perhaps the most important limitation revolves around concerns with misclassification of exposure, which is an inevitable limitation plaguing all environmental epidemiology research; it includes issues with spatial and temporal precision, as well as the time-activity and exposure environment of the individual participants. Epidemiologic research on long-term exposure to air pollution relies heavily on spatial variability amongst cohort participants to allow for a large enough difference in effect to be detectable.[29,97,98] In regard to spatial gradients, incorrectly-assigned exposure estimates based on spatial location may lead to misclassification bias of exposure. However, this is likely nondifferential and will bias modest effects, common to environmental exposure, towards the null rather than resulting in falsely positive associations.[99]

Daily time-activity was additionally not available, as was impact of residential mobility. It is well-known that people spend most of their time indoors; for instance, the findings of the National Human Activity Pattern Survey (NHAPS) suggested that people in the United States spend 87% of their time indoors.[100] However, people living in rural regions spend more time outdoors than their urban counterparts,[101] suggesting that the issue of nondifferential exposure misclassification bias due to time-activity patterns may be less prominent in rural populations compared to their urban counterparts. However, although in theory, neglecting to account for time-activity seems to be an important limitation, researchers have suggested that in application to observational studies, that is not the case; accounting for time-activity patterns does not affect how well ambient $PM_{2.5}$ concentrations approximate true personal exposure.[99]

Residential mobility refers to the movement of home location from one region with a certain annual $PM_{2.5}$ concentration, to a different region with a differing annual $PM_{2.5}$ concentration, leading to misclassification of exposure. A limitation of the current study is that data necessary to characterize residential mobility was not collected. Given the number of rural residents who choose a life of a

migrant worker in urban cities,[102] it is an understandable concern. However, one of the criteria for inclusion into the initial study was that eligible residents must not have plans to leave Deqing County, whether for change in work or home location. Moreover, residential mobility may not be a prominent issue in the present cohort because of their baseline demographics. Specifically, most migration in China occurs in early adulthood for better access to educational and career opportunities.[103] Nonetheless, residential mobility is an important source of bias in environmental epidemiology. Although misclassification of exposure is commonly non-differential, it may be differential if a person's health status and disease risk is influenced by whether they move or not.[104] Furthermore, research has shown that the bias due to residential mobility is likely minimal.[105]

Generalizability

Although residential mobility may not be an important limitation given the unique demographic of the study population, this uniqueness limits the generalizability to other populations. For instance, it is well-known that those who choose to become migrant workers tend to be healthier, younger, and have better cognitive function, as upward social mobility requires good health.[103] Additionally, rural residents who were once urban migrant workers often return to rural regions due to illness or family demands.[103] Therefore, those who remain in rural regions may therefore be more vulnerable than the general population.

This cohort study may not be generalizable to all rural regions of China, particularly with consideration for the difference in exposure related to living above or below the Huai River. China's Huai River Policy, which provides coal for indoor heating in cities north of the River for free or for heavily subsidized costs, has been shown to be associated with a 3-year difference in life expectancy for those living above compared to below the Huai River heating line.[106] Given that Deqing is a southern county

located below the Huai River, it is expected that the overall air pollution exposure will differ greatly from their northern counterparts. For instance, as was found in the HEI Report, domestic sources of ambient $PM_{2.5}$ in 2013 were the lowest sources in Zhejiang province, whereas the contribution of domestic sources of coal and biomass were much higher in northern provinces.[96] Nonetheless, the importance of representing differing cohorts with differing concentrations of exposure and underlying risks outweighs the issues of generalizability in this case.

3.6 Conclusion

In conclusion, association between long-term $PM_{2.5}$ exposure and risk of incident T2DM was examined in a rural cohort from Deqing County, China. The results indicate evidence of a relationship between $PM_{2.5}$ exposure and T2DM with potential effect modification by smoking status. $PM_{2.5}$ concentrations, exposure, and sources differ greatly depending on the region of study. The majority of air pollution epidemiologic research focuses on high-income urban regions in North America and Europe. The importance of these results suggest that even in rural China, long-term exposure to $PM_{2.5}$ increases the risk of rural residents developing T2DM. Further epidemiologic studies of the risk in other underrepresented rural regions and experimental studies on the biomechanisms are necessary to solidify the causality of the relationship between $PM_{2.5}$ and T2DM.

References for Article 1

1. Wang B, Xu D, Jing Z, Liu D, Yan S, Wang Y. Effect of long-term exposure to air pollution on type 2 diabetes mellitus risk: a systemic review and meta-analysis of cohort studies. Eur J Endocrinol. 2014 Nov;171(5):R173-182.

2. Rajagopalan S, Brook RD. Air pollution and type 2 diabetes: mechanistic insights. Diabetes. 2012 Dec;61(12):3037–45.

3. Thiering E, Heinrich J. Epidemiology of air pollution and diabetes. Trends Endocrinol Metab. 2015 Jul;26(7):384–94.

4. Dzhambov AM. Ambient air pollution and diabetes in China. Lancet Planet Health. 2018;2(2):e52–3.

5. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu NN, et al. The Lancet Commission on pollution and health. Lancet. 2018 03;391(10119):462–512.

6. Shao M, Tang X, Zhang Y, Li W. City clusters in China: air and surface water pollution. Front Ecol Environ. 2006;4(7):353–61.

7. Johnson TM, Guttikunda S, Wells GJ, Artaxo P, Bond TC, Russell AG, et al. Tools for improving air quality management : a review of top-down source apportionment techniques and their application in developing countries. The World Bank; 2011 Mar p. 1–220. Report No.: 81740.

8. World Health Organization. Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization; 2006.

9. Han L, Zhou W, Li W, Li L. Impact of urbanization level on urban air quality: a case of fine particles (PM(2.5)) in Chinese cities. Environ Pollut. 2014 Nov;194:163–70.

10. Lin G, Fu J, Jiang D, Hu W, Dong D, Huang Y, et al. Spatio-Temporal Variation of PM2.5 Concentrations and Their Relationship with Geographic and Socioeconomic Factors in China. Int J Environ Res Public Health. 2014 Jan;11(1):173–86.

11. Brook RD, Newby DE, Rajagopalan S. Air Pollution and Cardiometabolic Disease: An Update and Call for Clinical Trials. Am J Hypertens. 2017 Dec 8;31(1):1–10.

12. Polichetti G, de Sio A, Fiore E, Polichetti R. The relationship between ambient air pollution and diabetes: A brief literature review. In 2015.

13. He D, Wu S, Zhao H, Qiu H, Fu Y, Li X, et al. Association between particulate matter 2.5 and diabetes mellitus: A meta-analysis of cohort studies. J Diabetes Investig. 2017 Sep;8(5):687–96.

14. Esposito K, Petrizzo M, Maiorino MI, Bellastella G, Giugliano D. Particulate matter pollutants and risk of type 2 diabetes: a time for concern? Endocrine. 2016 Jan;51(1):32–7.

15. Yang B-Y, Qian ZM, Li S, Chen G, Bloom MS, Elliott M, et al. Ambient air pollution in relation to diabetes and glucose-homoeostasis markers in China: a cross-sectional study with findings from the 33 Communities Chinese Health Study. Lancet Planet Health. 2018;2(2):e64–73.

16. Fu C, Chen Y, Wang F, Wang X, Song J, Jiang Q. High prevalence of hyperglycaemia and the impact of high household income in transforming rural China. BMC Public Health. 2011 Nov 12;11:862.

17. Chen X, Wu Z, Chen Y, Wang X, Zhu J, Wang N, et al. Risk score model of type 2 diabetes prediction for rural Chinese adults: the Rural Deqing Cohort Study. J Endocrinol Invest. 2017 Oct;40(10):1115–23.

18. Fu C, Chen Y, Wang F, Wang X, Wang W, Zhao G, et al. Chronic Diseases in Adults Living in Rural Communities of China. In 2012.

19. van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC, et al. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environ Sci Technol. 2016 Apr 5;50(7):3762–72.

20. Boys BL, Martin RV, van Donkelaar A, MacDonell RJ, Hsu NC, Cooper MJ, et al. Fifteen-Year Global Time Series of Satellite-Derived Fine Particulate Matter. Environ Sci Technol. 2014 Oct 7;48(19):11109–18.

21. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. Environ Health Perspect. 2015 Feb;123(2):135–43.

22. Motulsky H. Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking. New York, NY: Oxford University Press; 1995.

23. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet. 2004 Jan 10;363(9403):157–63.

24. Yi H, Zhang L, Luo R, Shi Y, Mo D, Chen X, et al. Dropping out: Why are students leaving junior high in China's poor rural areas? Int J Educ Dev. 2012 Jul 1;32(4):555–63.

25. Qiu H, Schooling CM, Sun S, Tsang H, Yang Y, Lee RS-Y, et al. Long-term exposure to fine particulate matter air pollution and type 2 diabetes mellitus in elderly: A cohort study in Hong Kong. Environ Int. 2018;113:350–6.

26. Lemeshow S, Hosmer D, May S. Applied Survival Analysis: Regression Modeling of Time-to-Event Data. 2nd ed. Hoboken, New Jersey: Wiley-Interscience; 2008.

27. Gelfand AE, Fuentes M, Hoeting JA, Smith RL, Fuentes M, Hoeting JA, et al. Handbook of Environmental and Ecological Statistics. Chapman and Hall/CRC; 2019.

28. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health. 1999;20:145–57.

29. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation. 2010 Jun 1;121(21):2331–78.

30. Lipsett MJ, Ostro BD, Reynolds P, Goldberg D, Hertz A, Jerrett M, et al. Long-Term Exposure to Air Pollution and Cardiorespiratory Disease in the California Teachers Study Cohort. Am J Respir Crit Care Med. 2011 Oct 1;184(7):828–35.

31. Chen H, Burnett RT, Kwong JC, Villeneuve PJ, Goldberg MS, Brook RD, et al. Risk of incident diabetes in relation to long-term exposure to fine particulate matter in Ontario, Canada. Environ Health Perspect. 2013 Jul;121(7):804–10.

32. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. Int J Epidemiol. 2013 Aug;42(4):1187–95.

33. Cao J, Yang C, Li J, Chen R, Chen B, Gu D, et al. Association between long-term exposure to outdoor air pollution and mortality in China: a cohort study. J Hazard Mater. 2011 Feb 28;186(2–3):1594–600.

34. Lao XQ, Guo C, Chang L-Y, Bo Y, Zhang Z, Chuang YC, et al. Long-term exposure to ambient fine particulate matter (PM2.5) and incident type 2 diabetes: a longitudinal cohort study. Diabetologia. 2019;62(5):759–69.

35. Strak M, Janssen N, Beelen R, Schmitz O, Vaartjes I, Karssenberg D, et al. Long-term exposure to particulate matter, NO2 and the oxidative potential of particulates and diabetes prevalence in a large national health survey. Environ Int. 2017;108:228–36.

36. Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. Ambient PM2.5, O_3, and NO_2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ Health Perspect. 2015 Nov;123(11):1180–6.

37. Chen H, Goldberg MS, Burnett RT, Jerrett M, Wheeler AJ, Villeneuve PJ. Long-term exposure to traffic-related air pollution and cardiovascular mortality. Epidemiology. 2013 Jan;24(1):35–43.

38. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. Ann Statist. 1982 Dec;10(4):1100–20.

39. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press; 2003.

40. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag; 2000. (Statistics for Biology and Health).

41. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. 2nd ed. New York: Springer-Verlag; 2012. (Statistics for Biology and Health).

42. Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd ed. Springer International Publishing; 2015. (Springer Series in Statistics).

43. Thomas D. Statistical Methods in Environmental Epidemiology. New York, USA: Oxford University Press; 2009. 352 p.

44. Weber MB, Oza-Frank R, Staimez LR, Ali MK, Narayan KMV. Type 2 diabetes in Asians: prevalence, risk factors, and effectiveness of behavioral intervention at individual and population levels. Annu Rev Nutr. 2012 Aug 21;32:417–39.

45. Chan JCN, Malik V, Jia W, Kadowaki T, Yajnik CS, Yoon K-H, et al. Diabetes in Asia: Epidemiology, Risk Factors, and Pathophysiology. JAMA. 2009 May 27;301(20):2129–40.

46. Zhang N, Du SM, Ma GS. Current lifestyle factors that increase risk of T2DM in China. Eur J Clin Nutr. 2017 Jul;71(7):832–8.

47. Honda T, Pun VC, Manjourides J, Suh H. Associations between long-term exposure to air pollution, glycosylated hemoglobin and diabetes. Int J Hyg Environ Health. 2017;220(7):1124–32.

48. Brook RD, Cakmak S, Turner MC, Brook JR, Crouse DL, Peters PA, et al. Long-term fine particulate matter exposure and mortality from diabetes in Canada. Diabetes Care. 2013 Oct;36(10):3313–20.

49. Paul LA, Burnett RT, Kwong JC, Hystad P, van Donkelaar A, Bai L, et al. The impact of air pollution on the incidence of diabetes and survival among prevalent diabetes cases. Environ Int. 2020 Jan;134:105333.

50. Park SK, Adar SD, O'Neill MS, Auchincloss AH, Szpiro A, Bertoni AG, et al. Long-term exposure to air pollution and type 2 diabetes mellitus in a multiethnic cohort. Am J Epidemiol. 2015 Mar 1;181(5):327–36.

51. Hansen AB, Ravnskjær L, Loft S, Andersen KK, Bräuner EV, Baastrup R, et al. Long-term exposure to fine particulate matter and incidence of diabetes in the Danish Nurse Cohort. Environ Int. 2016 May;91:243–50.

52. Renzi M, Cerza F, Gariazzo C, Agabiti N, Cascini S, Di Domenicantonio R, et al. Air pollution and occurrence of type 2 diabetes in a large cohort study. Environ Int. 2018;112:68–76.

53. Weinmayr G, Hennig F, Fuks K, Nonnemacher M, Jakobs H, Möhlenkamp S, et al. Long-term exposure to fine particulate matter and incidence of type 2 diabetes mellitus in a cohort study: effects of total and traffic-specific air pollution. Environ Health. 2015 Jun 19;14:53.

54. Liang F, Yang X, Liu F, Li J, Xiao Q, Chen J, et al. Long-term exposure to ambient fine particulate matter and incidence of diabetes in China: A cohort study. Environ Int. 2019;126:568–75.

55. Li C-Y, Wu C-D, Pan W-C, Chen Y-C, Su H-J. Association Between Long-term Exposure to PM2.5 and Incidence of Type 2 Diabetes in Taiwan: A National Retrospective Cohort Study. Epidemiology. 2019 Jul;30 Suppl 1:S67–75.

56. Pope CA, Turner MC, Burnett RT, Jerrett M, Gapstur SM, Diver WR, et al. Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality. Circ Res. 2015 Jan 2;116(1):108–15.

57. Coogan PF, White LF, Yu J, Burnett RT, Seto E, Brook RD, et al. PM2.5 and Diabetes and Hypertension Incidence in the Black Women's Health Study. Epidemiology. 2016 Mar;27(2):202–10.

58. Hagger-Johnson G. Introduction to Research Methods and Data Analysis in the Health Sciences [Internet]. Routledge; 2014 [cited 2020 Jun 15]. Available from: https://www.taylorfrancis.com/books/9781315771113

59. Yin P, Brauer M, Cohen A, Burnett RT, Liu J, Liu Y, et al. Long-term Fine Particulate Matter Exposure and Nonaccidental and Cause-specific Mortality in a Large National Cohort of Chinese Men. Environ Health Perspect. 2017 07;125(11):117002.

60. Jin C, Cheng J, Lu Y, Huang Z, Cao F. Spatial inequity in access to healthcare facilities at a county level in a developing country: a case study of Deqing County, Zhejiang, China. Int J Equity Health. 2015 Aug 19;14:67.

61. Qiu Y, Zhao X, Fan D, Li S. Geospatial Disaggregation of Population Data in Supporting SDG Assessments: A Case Study from Deqing County, China. IJGI. 2019 Aug;8(8):356.

62. Wang QW, Fan DQ, Qiu Y, Miao YQ. Measuring Regional Progress Towards SDG 3 With Geospatial And Statistical Information – A Case Study Of Deqing County. In: ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Copernicus GmbH; 2019. p. 89–95.

63. Lin G, Fu J, Jiang D, Hu W, Dong D, Huang Y, et al. Spatio-Temporal Variation of PM2.5 Concentrations and Their Relationship with Geographic and Socioeconomic Factors in China. Int J Environ Res Public Health. 2014 Jan;11(1):173–86.

64. Chen X, Zhang L-W, Huang J-J, Song F-J, Zhang L-P, Qian Z-M, et al. Long-term exposure to urban air pollution and lung cancer mortality: A 12-year cohort study in Northern China. Sci Total Environ. 2016 Nov 15;571:855–61.

65. Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. Particulate Matter Air Pollution and the Risk of Incident CKD and Progression to ESRD. J Am Soc Nephrol. 2018;29(1):218–30.

66. Xu Y, Wang L, He J, Bi Y, Li M, Wang T, et al. Prevalence and control of diabetes in Chinese adults. JAMA. 2013 Sep 4;310(9):948–59.

67. Zhou M, Astell-Burt T, Bi Y, Feng X, Jiang Y, Li Y, et al. Geographical variation in diabetes prevalence and detection in china: multilevel spatial analysis of 98,058 adults. Diabetes Care. 2015 Jan;38(1):72–81.

68. Bragg F, Holmes MV, Iona A, Guo Y, Du H, Chen Y, et al. Association Between Diabetes and Cause-Specific Mortality in Rural and Urban Areas of China. JAMA. 2017 17;317(3):280–9.

69. Ma RCW. Epidemiology of diabetes and diabetic complications in China. Diabetologia. 2018;61(6):1249–60.

70. Hu D, Xie J, Fu P, Zhou J, Yu D, Whelton PK, et al. Central rather than overall obesity is related to diabetes in the Chinese population: the InterASIA study. Obesity. 2007 Nov;15(11):2809–16.

71. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. 2nd ed. New York: Springer-Verlag; 2005. (Statistics for Biology and Health).

72. Collett D. Modelling Survival Data in Medical Research. 2nd ed. Bristol, UK: Chapman and Hall/CRC; 2003.

73. Martinussen T, Scheike TH. Dynamic Regression Models for Survival Data. New York: Springer-Verlag; 2006. (Statistics for Biology and Health).

74. Liu F, Chen G, Huo W, Wang C, Liu S, Li N, et al. Associations between long-term exposure to ambient air pollution and risk of type 2 diabetes mellitus: A systematic review and meta-analysis. Environ Pollut. 2019 Sep 1;252:1235–45.

75. Yang M, Cheng H, Shen C, Liu J, Zhang H, Cao J, et al. Effects of long-term exposure to air pollution on the incidence of type 2 diabetes mellitus: a meta-analysis of cohort studies. Environ Sci Pollut Res Int. 2019 Dec 6;

76. Yang B-Y, Fan S, Thiering E, Seissler J, Nowak D, Dong G-H, et al. Ambient air pollution and diabetes: A systematic review and meta-analysis. Environ Res. 2020 Jan 1;180:108817.

77. Lim CC, Hayes RB, Ahn J, Shao Y, Silverman DT, Jones RR, et al. Association between long-term exposure to ambient air pollution and diabetes mortality in the US. Environ Res. 2018;165:330–6.

78. Lucht S, Hennig F, Moebus S, Ohlwein S, Herder C, Kowall B, et al. All-source and source-specific air pollution and 10-year diabetes Incidence: Total effect and mediation analyses in the Heinz Nixdorf recall study. Environ Int. 2020 Mar;136:105493.

79. Rao X, Patel P, Puett R, Rajagopalan S. Air pollution as a risk factor for type 2 diabetes. Toxicol Sci. 2015 Feb;143(2):231–41.

80. Liu C, Ying Z, Harkema J, Sun Q, Rajagopalan S. Epidemiological and experimental links between air pollution and type 2 diabetes. Toxicol Pathol. 2013 Feb;41(2):361–73.

81. Rao X, Montresor-Lopez J, Puett R, Rajagopalan S, Brook RD. Ambient air pollution: an emerging risk factor for diabetes mellitus. Curr Diab Rep. 2015 Jun;15(6):603.

82. Heck TG, Fiorin PBG, Frizzo MN, Ludwig MS. Fine Particulate Matter (PM2.5) Air Pollution and Type 2 Diabetes Mellitus (T2DM): When Experimental Data Explains Epidemiological Facts. In: G. AR, editor. Diabetes and Its Complications. InTech; 2018.

83. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. PLoS ONE. 2018;13(3):e0194127.

84. Crouse DL, Erickson AC, Christidis T, Pinault L, van Donkelaar A, Li C, et al. Evaluating the Sensitivity of PM2.5-Mortality Associations to the Spatial and Temporal Scale of Exposure Assessment. Epidemiology. 2019 Nov 4;

85. White LF, Yu J, Jerrett M, Coogan P. Temporal aspects of air pollutant measures in epidemiologic analysis: a simulation study. Sci Rep. 2016 Jan 21;6:19691.

86. Lefler JS, Higbee JD, Burnett RT, Ezzati M, Coleman NC, Mann DD, et al. Air pollution and mortality in a large, representative U.S. cohort: multiple-pollutant analyses, and spatial and temporal decompositions. Environ Health. 2019 21;18(1):101.

87. Ma Z, Hu X, Sayer AM, Levy R, Zhang Q, Xue Y, et al. Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004–2013. Environ Health Perspect. 2016 Feb;124(2):184–92.

88. Rohde RA, Muller RA. Air Pollution in China: Mapping of Concentrations and Sources. PLoS ONE. 2015;10(8):e0135749.

89. ADA. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. Diabetes Care. 2020 Jan;43(Suppl 1):S14–31.

90. Jia W, Weng J, Zhu D, Ji L, Lu J, Zhou Z, et al. Standards of medical care for type 2 diabetes in China 2019. Diabetes Metab Res Rev. 2019;35(6):e3158.

91. Diabetes Canada Clinical Practice Guidelines Expert Committee, Punthakee Z, Goldenberg R, Katz P. Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. Can J Diabetes. 2018 Apr;42 Suppl 1:S10–5.

92. Sacks DB. A1C versus glucose testing: a comparison. Diabetes Care. 2011 Feb;34(2):518–23.

93. Yang T, Rockett IRH, Wu Y, Chen WW, Lu Z. Agreement among multiple measures of self-reported smoking status in Chinese urban residents. Public Health. 2010 Aug;124(8):437–43.

94. Wang L, Gao P, Zhang M, Huang Z, Zhang D, Deng Q, et al. Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. JAMA. 2017 27;317(24):2515–23.

95. Green MA. Do we need to think beyond BMI for estimating population-level health risks? J Public Health. 2016 Mar 1;38(1):192–3.

96. GBD Maps Working Group. Burden of Disease Attributable to Coal-Burning and Other Air Pollution Sources in China. Boston, MA: Health Effects Institute; 2016. Report No.: 20.

97. WHO Global Air Quality Guidelines Working Group on Risk of Bias Assessment. Risk of bias assessment instrument for systematic reviews informing WHO global air quality guidelines (2020). Copenhagen, Denmark: World Health Organization Regional Office for Europe; 2020 Feb.

98. Brauer M. How much, how long, what, and where: air pollution exposure assessment for epidemiologic studies of respiratory disease. Proc Am Thorac Soc. 2010 May;7(2):111–5.

99. Avery CL, Mills KT, Williams R, McGraw KA, Poole C, Smith RL, et al. Estimating Error in Using Residential Outdoor PM2.5 Concentrations as Proxies for Personal Exposures: A Meta-analysis. Environ Health Perspect. 2010 May;118(5):673–8.

100. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Expo Anal Environ Epidemiol. 2001 Jun;11(3):231–52.

101. Lü J, Liang L, Feng Y, Li R, Liu Y. Air Pollution Exposure and Physical Activity in China: Current Knowledge, Public Health Implications, and Future Research Needs. Int J Environ Res Public Health. 2015 Nov 20;12(11):14887–97.

102. Chan KW. China: internal migration. In: The Encyclopedia of Global Human Migration. American Cancer Society; 2013.

103. Xu H, Dupre ME, Østbye T, Vorderstrasse AA, Wu B. Residential Mobility and Cognitive Function Among Middle-Aged and Older Adults in China. Res Aging. 2019;41(1):3–30.

104. Pendakur R, Young N. Putting on the moves: Individual, household, and community-level determinants of residential mobility in Canada. Demographic Research. 2013 Oct 10;29(28):767–96.

105. Oudin A, Forsberg B, Strömgren M, Beelen R, Modig L. Impact of Residential Mobility on Exposure Assessment in Longitudinal Air Pollution Studies: A Sensitivity Analysis within the ESCAPE Project [Internet]. Vol. 2012, The Scientific World Journal. Hindawi; 2012 [cited 2020 Aug 27]. p. e125818. Available from: https://www.hindawi.com/journals/tswj/2012/125818/

106. Ebenstein A, Fan M, Greenstone M, He G, Zhou M. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proc Natl Acad Sci USA. 2017 26;114(39):10384–9.

107. Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med. 2007 Feb 1;356(5):447–58.

108. Hernán MA. The hazards of hazard ratios. Epidemiology. 2010 Jan;21(1):13–5.

109. Allison PD. Survival Analysis Using SAS: A Practical Guide. 2nd ed. North Carolina, USA: SAS Institute Inc.; 2010.

Supplementary Material for Article 1

Supplementary Material 1. Assignment of Exposure Metrics

The main objective was to assess the relationship between long-term exposure to ambient fine particulate matter ($PM_{2.5}$) and incidence of type 2 diabetes mellitus (T2DM), based on data from a cohort study conducted in Deqing County, Zhejiang Province, China. By linking participants' address codes at study enrolment with the corresponding satellite imaging region, each participant was assigned multiple annual mean concentrations of $PM_{2.5}$ and its constituents for the period of 2006–2014. It was assumed that the mean concentrations of $PM_{2.5}$ at a participant's place of residence at study enrolment were representative of his or her overall personal exposure. For modelling purposes, subject-level $PM_{2.5}$ exposures were created using three methods.

In Method 1, participants were assigned the annual concentration of $PM_{2.5}$ and its constituents from the year of enrolment; this is the time-fixed, baseline exposure. In Method 2, participants were assigned the 3-year average annual concentration of $PM_{2.5}$ at year of entry; this is the time-fixed, 3-year average-assigned exposure. In Method 3, participants were assigned the annual concentration for every year they were enrolled in the study, reflecting changes in concentration over the years; this is the time-varying annual exposure. For time-varying exposure, the dataset was reshaped to fit a counting process data structure, in which each participant had multiple rows and each row represented a risk set time interval at which a participant's exposure changed, for which we have data on. Each row of data represented a subsequent year that participants were followed in the study.

For clarity, theoretical participants are presented below, with examples to demonstrate how each method was used to create exposure assignments.

Table S1.1. Data: Satellite-Derived Data on Ambient Fine Particulate Matter Concentrations

Participant ID	Entry year	Exit year	Event	Average annual concentration of $PM_{2.5}$, corresponding to personal address code								
				2006	2007	2008	2009	2010	2011	2012	2013	2014
1	2006	2008	1	49	48	48	46	50	51	53	54	55
2	2008	2012	0	40	41	42	42	44	45	44	46	47

Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$.

Table S1.2. Main Method 1: Time-Fixed, Baseline Exposure to Ambient Fine Particulate Matter

Participant ID	Years in cohort	Person-years	Event at end of follow-up	Participant exposure	
1	2006–2009	3	1	**[$PM_{2.5}$] in** 2006:	49
2	2008–2012	5	0	**[$PM_{2.5}$] in** 2008:	42

Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$; [$PM_{2.5}$], concentration of ambient fine particulate matter.

One-year annual concentration of $PM_{2.5}$ in the year the participant entered the study cohort is used as a proxy for "true" long-term personal exposure to $PM_{2.5}$. Although time-varying exposure is logically considered the more accurate exposure assignment method, as it considered the changes over time, there are situations for which baseline exposure can be the more appropriate method.[64] *In other cases, a single year may better approximate personal exposure due to concentration measurement; for example, in a year that a substantial increase in the number of network monitors occurred in the one year, thus that year is used as a proxy for long-term exposure.*[107] *Furthermore, previous studies have shown that baseline exposure are valid proxy measures for long-term exposure.*[37]

Table S1.3. Main Method 2: Time-Fixed, Three-Year Average Exposure to Ambient Fine Particulate Matter Since Baseline

Participant ID	Years in cohort	Person-years	Event at end of follow-up	Participant exposure	
1	2006–2009	3	1	**average [$PM_{2.5}$] in** 2006–2008:	48.3
2	2008–2012	5	0	**average [$PM_{2.5}$] in** 2008–2010:	43.3

Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$; [$PM_{2.5}$], concentration of ambient fine particulate matter.

Average of three annual concentrations of $PM_{2.5}$ since the year the participant entered the study cohort is used as a proxy for "true" long-term personal exposure to $PM_{2.5}$. The purpose of a time-fixed, three-year average exposure, compared to a time-fixed, baseline exposure, is to stabilize estimates that may be effected by noise in annual satellite-derived values.[36]

Table S1.4. Main Method 3: Yearly Time-Varying Exposure to Ambient Fine Particulate Matter

Participant ID	Years in cohort	Person-years	Risk set (Star of year	End of year]	Event at end of each risk set	Participant exposure	
1	2006	3	0	1	0	**[$PM_{2.5}$] in** 2006:	49
	2007		1	2	0	**[$PM_{2.5}$] in** 2007:	48
	2008		2	3	1	**[$PM_{2.5}$] in** 2008:	48
2	2008	5	0	1	0	**[$PM_{2.5}$] in** 2008:	42
	2009		1	2	0	**[$PM_{2.5}$] in** 2009:	42
	2010		2	3	0	**[$PM_{2.5}$] in** 2010:	44
	2011		3	4	0	**[$PM_{2.5}$] in** 2011:	45
	2012		4	5	0	**[$PM_{2.5}$] in** 2012:	44

Half-open interval, i.e. (Start of year, End of Year], refers to the start (exclusive) to end of year (inclusive), with the purpose of insuring no time overlap between risk set intervals.
Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of µg/m^3; [$PM_{2.5}$], concentration of ambient fine particulate matter.
One-year annual concentration of $PM_{2.5}$ for every year the participant is contributing to the study cohort risk set, i.e. from the year the participant entered the study cohort to the year the participant developed the event, was lost to follow-up, or the end of the study, whichever came first. Given that yearly time-varying exposure accounts for time-dependent changes in exposure, this method of exposure assignment more accurately depicts "true" long-term personal exposure to $PM_{2.5}$.[28,108]

Table S1.5. Sensitivity Analysis Method 4: Time-Fixed, Full Study Period (2006–2014) Average Exposure to Ambient Fine Particulate Matter

Participant ID	Years in cohort	Person-years	Event at end of follow-up	Participant exposure	
1	2006–2009	3	1	**average [$PM_{2.5}$] in** 2006–2014:	50.4
2	2008–2012	5	0	**average [$PM_{2.5}$] in** 2006–2014:	43.4

Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of µg/m^3; [$PM_{2.5}$], concentration of ambient fine particulate matter.
The average of nine annual concentrations of $PM_{2.5}$ for the full study period is used as a proxy for "true" long-term personal exposure to $PM_{2.5}$. This sensitivity analysis serves a similar purpose to three-year average exposure, in that the averaging is expected to provide stability to estimates.[21,31]*The aim of this sensitivity analysis is to affirm whether the main methods, no or shorter averaging, lead to effect estimates that are not replicated when using longer averaging.*

Table S1.6. Sensitivity Analysis Method 5: Three-Year Moving Average Exposure to Ambient Fine Particulate Matter

Participant ID	Years in cohort	Person-years	Risk set		Event at end of each risk set	Participant exposure
			(Star of year	End of year]		
1	2006	3	0	1	0	**average [$PM_{2.5}$] in** 2004–2006:
	2007		1	2	0	**average [$PM_{2.5}$] in** 2005–2007:
	2008		2	3	1	**average [$PM_{2.5}$] in** 2006–2008:

Half-open interval, i.e. (Start of year, End of Year], refers to the start (exclusive) to end of year (inclusive), with the purpose of insuring no time overlap between risk set intervals.
Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$; [$PM_{2.5}$], concentration of ambient fine particulate matter.
The average of three annual concentrations of $PM_{2.5}$ are provided for every year the participant is contributing to the study cohort risk set, i.e. from the year the participant entered the study cohort to the year the participant developed the event, was lost to follow-up, or the end of the study, whichever came first. As previously mentioned, averaging is expected to stabilize satellite-derived estimates of PM2.5 that may vary year to year due to instrumental error.[21] As a exposure assignment method that varies by time, this three-year moving average serves as a sensitivity analysis for the yearly time-varying exposure, to assess whether averaging instead of one-year time-varying exposure leads to attenuated effects. Additionally, this three-year moving average also serves as a sensitivity analysis for the three-year average exposure, to assess whether accounting for change in exposure over time will effect the results.

Table S1.7. Sensitivity Analysis Method 6: Three-Year Moving Average Exposure with One-Year Lag to Ambient Fine Particulate Matter

Participant ID	Years in cohort	Person-years	Risk set		Event at end of each risk set	Participant exposure
			(Star of year	End of year]		
1	2006	3	0	1	0	**average [$PM_{2.5}$] in** 2003–2005:
	2007		1	2	0	**average [$PM_{2.5}$] in** 2004–2006:
	2008		2	3	1	**average [$PM_{2.5}$] in** 2005–2007:

Half-open interval, i.e. (Start of year, End of Year], refers to the start (exclusive) to end of year (inclusive), with the purpose of insuring no time overlap between risk set intervals.
Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$; [$PM_{2.5}$], concentration of ambient fine particulate matter.
The average of three annual concentrations of $PM_{2.5}$ are provided for every year the participant is contributing to the study cohort risk set, as an average of the three years prior to each assigned year. A lag ensures temporality needed to infer causality.[26,109]

Table S1.8. Sensitivity Analysis Method 7: Yearly Time-Varying Exposure with One-Year Lag to Ambient Fine Particulate Matter

Participant ID	Years in cohort	Person-years	Risk set		Event at end of each risk set	Participant exposure
			(Star of year	End of year]		
1	2006	3	0	1	0	**[$PM_{2.5}$] in** 2005:
	2007		1	2	0	**[$PM_{2.5}$] in** 2006:
	2008		2	3	1	**[$PM_{2.5}$] in** 2007:

Half-open interval, i.e. (Start of year, End of Year], refers to the start (exclusive) to end of year (inclusive), with the purpose of insuring no time overlap between risk set intervals.
Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of $\mu g/m^3$; [$PM_{2.5}$], concentration of ambient fine particulate matter.
One-year annual concentration of $PM_{2.5}$ for every year the participant is contributing to the study cohort risk set, with the concentration assigned as the year prior. Similar to method 6, this yearly time-varying exposure with one-year lag serves as a sensitivity analysis for the main method 3, yearly time-varying exposure, to assess whether the a lag period effects the results.

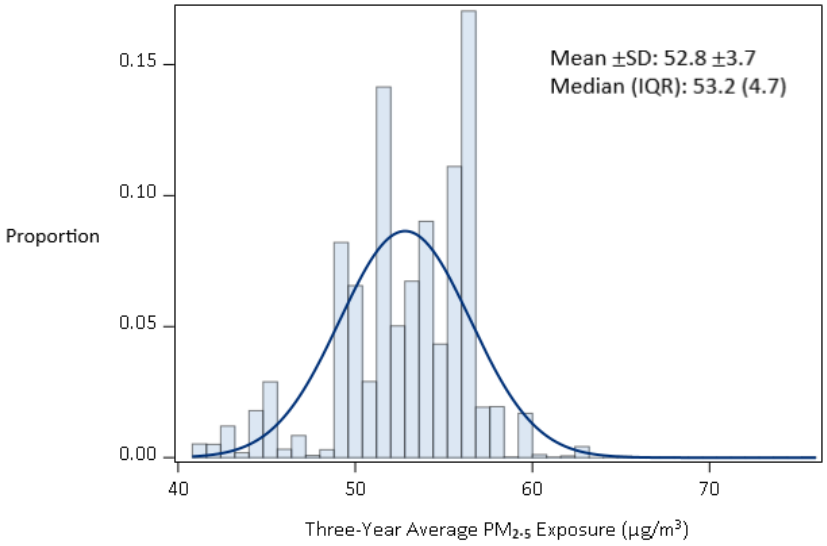
Supplementary Material 2. Three-Year Average $PM_{2.5}$

Table S2.1. Descriptive Statistics for Time-Fixed, Three-Year Average Concentration of Ambient Fine Particulate Matter and Its Constituents in Deqing County

Three-year average concentration (μg/m³)	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
$PM_{2.5}$	52.82	3.69	40.86	50.94	53.22	55.67	75.61	4.72

Abbreviations: SD, standard deviation; IQR, interquartile range; $PM_{2.5}$, ambient fine particulate matter.

Figure S2.1. Histogram of Three-Year Average Ambient Fine Particulate Matter Exposure in Deqing County, from 2006 to 2014



Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of μg/m³; SD, standard deviation; IQR, interquartile range.

Table S2.2. Effect Modification of Time-Fixed, Three-Year Average Exposure to Ambient Fine Particulate Matter and Selected Baseline Characteristics on Incidence of Type 2 Diabetes Mellitus

Model	Variables in three-year average exposure model	β	SE	p-value for interaction
1	**Unadjusted model**: $PM_{2.5}$	—	—	—
2	$PM_{2.5}$ + Gender + ($PM_{2.5}$ × Gender)			0.0884
	Male vs. Female	0.0466	0.0274	
3	$PM_{2.5}$ + Age + ($PM_{2.5}$ × Age)			0.0720
	<40 vs. 50–59	-0.2068	0.0820	
	40–49 vs. 50–59	-0.0623	0.0367	

Model	Variables in three-year average exposure model	β	SE	p-value for interaction
	60–69 vs. 50–59	-0.0074	0.0361	
	≥70 vs. 50–59	-0.0181	0.0433	
4	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0136
	Ever smoked vs. Never smoked	0.0774	0.0314	
5	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)			0.0045
	Overweight vs. Normal or underweight	-0.06776	0.02819	
	Obese vs. Normal or underweight	-0.16138	0.06086	

A p-value for interaction of <0.05 was considered statistically significant and an indicator of potential effect modification. Reference categories were chosen based on category with largest sample size in order to ensure stable results. All covariates presented as potential effect modifiers were selected a priori based on the current literature.
*Abbreviations: β, regression coefficient for $PM_{2.5}$; SE, standard error; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (*kg/m^2*), categorized as normal or underweight (<24.9), overweight (25.0–29.9), and obese (≥30).*

Table S2.3. Likelihood Ratio Tests for the Inclusion of Two Effect Modifiers of the Association Between Incidence of Type 2 Diabetes Mellitus and Three-Year Average Exposure to Ambient Fine Particulate Matter

Model	Variables in three-year average exposure model	-2 LOG L	*df*	LR test	p-value
1	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)	7,418.501	5	—	—
2	$PM_{2.5}$ + BMI + smoking status + ($PM_{2.5}$ × BMI) + ($PM_{2.5}$ × smoking status)	7,411.198	7	7.303	0.02595

A p-value for the likelihood ratio test of <0.05 was considered statistically significant and an indicator of better model fit. Although the p-value for including both effect modifiers, BMI and smoking status, given the small sample size per individual category, they are not presented in the analytic results presented in Table 10; instead, they are presented separately and this limitation will be described in the discussion section.
*Abbreviations: -2 LOG L, -2 log-likelihood (deviance); df, degrees of freedom; LR test, likelihood ratio chi–square test; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (*kg/m^2*), categorized as normal or underweight (<24.9), overweight (25.0–29.9), and obese (≥30).*

Table S2.4. Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Adjusted Independently By Potential Confounders

Model	Variables in three-year average exposure model	HR	95% CI	p-value	β	Percent change
1	**Unadjusted model**: $PM_{2.5}$	1.27	0.98, 1.65	0.0747	0.02392	—
2	$PM_{2.5}$ + Gender	1.27	0.97, 1.65	0.0797	0.02358	1.44
3	$PM_{2.5}$ + Age	1.35	1.03, 1.77	0.0306	0.02992	20.05
4	$PM_{2.5}$ + BMI	1.27	0.98, 1.66	0.0756	0.02401	0.37
5	$PM_{2.5}$ + Exercise	1.28	0.98, 1.66	0.0723	0.02427	1.44
6	$PM_{2.5}$ + Smoking status	1.27	0.98, 1.65	0.0762	0.02390	0.08

Model	Variables in three-year average exposure model	HR	95% CI	p-value	β	Percent change
7	$PM_{2.5}$ + Passive smoking exposure	1.31	1.01, 1.71	0.0445	0.02728	12.32
8	$PM_{2.5}$ + Vegetable consumption	1.37	1.04, 1.79	0.0231	0.03122	23.38
9	$PM_{2.5}$ + Fruit consumption	1.26	0.97, 1.65	0.0895	0.02323	2.97
10	$PM_{2.5}$ + Family history of diabetes	1.30	1.00, 1.68	0.0520	0.02596	7.86
11	$PM_{2.5}$ + History of disease	1.31	1.01, 1.70	0.0429	0.02701	11.44
12	$PM_{2.5}$ + Education	1.26	0.96, 1.64	0.0930	0.02289	4.50
13	$PM_{2.5}$ + Occupation	1.33	1.02, 1.73	0.0330	0.02860	16.36
14	$PM_{2.5}$ + Family income	1.27	0.98, 1.65	0.0771	0.02374	0.76
15	**Parsimoniously-adjusted model**[a]	1.45	1.10, 1.92	0.0082	—	—

Percent change of regression *coefficient was calculated as |(nested model* β *– fuller model* β*) ÷ fuller model* β*|×100%. A percent change of* regression *coefficient ≥10% was considered strong evidence of confounding of the association between long-term exposure to ambient fine particulate matter and incidence of type 2 diabetes mellitus. All covariates presented as potential confounders were selected a priori based on the current literature. the parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%.*

[a]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, passive smoking status, vegetable consumption, history of disease, and occupation.

Abbreviations: HR, hazard ratio, per 10-μg/m³ increase in $PM_{2.5}$; CI, confidence interval; β*, regression coefficient; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index.*

Table S2.5. Effect Modification of Three-Year Average Exposure to Ambient Fine Particulate Matter in Parsimoniously- and Fully-Adjusted Models

Model	Variables in baseline exposure model	β	SE	p-value for interaction
1	**Parsimoniously-adjusted model**[a]	—	—	—
2	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0242
	Ever smoked vs. Never smoked	0.06987	0.03100	
3	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)			0.0038
	Overweight vs. Normal or underweight	-0.07156	0.02844	
	Obese vs. Normal or underweight	-0.16453	0.06321	
4	**Fully-adjusted model**[b]			
5	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0223
	Ever smoked vs. Never smoked	0.07169	0.03137	
6	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)			0.0038
	Overweight vs. Normal or underweight	-0.07266	0.02871	
	Obese vs. Normal or underweight	-0.16360	0.06297	

[a]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, passive smoking status, vegetable consumption, history of disease, and occupation.

[b]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Table S2.6. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Smoking Status

Variables in three-year average exposure model	Overall		Smoking Status			
			Non-smokers		Smokers	
	(Total cases = 386)		(No. of cases = 308)		(No. of cases = 78)	
	HR	95% CI	HR	95% CI	HR	95% CI
Unadjusted model[a]	1.27	0.98, 1.65	1.06	0.79, 1.42	2.36	1.37, 4.09
Parsimoniously-adjusted model[b]	1.45	1.10, 1.92	1.24	0.91, 1.70	2.59	1.46, 4.58
Fully-adjusted model[c]	1.43	1.07, 1.89	1.21	0.88, 1.66	2.66	1.48, 4.81

The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%. the fully-adjusted model includes all covariates chosen a priori based on the current literature.

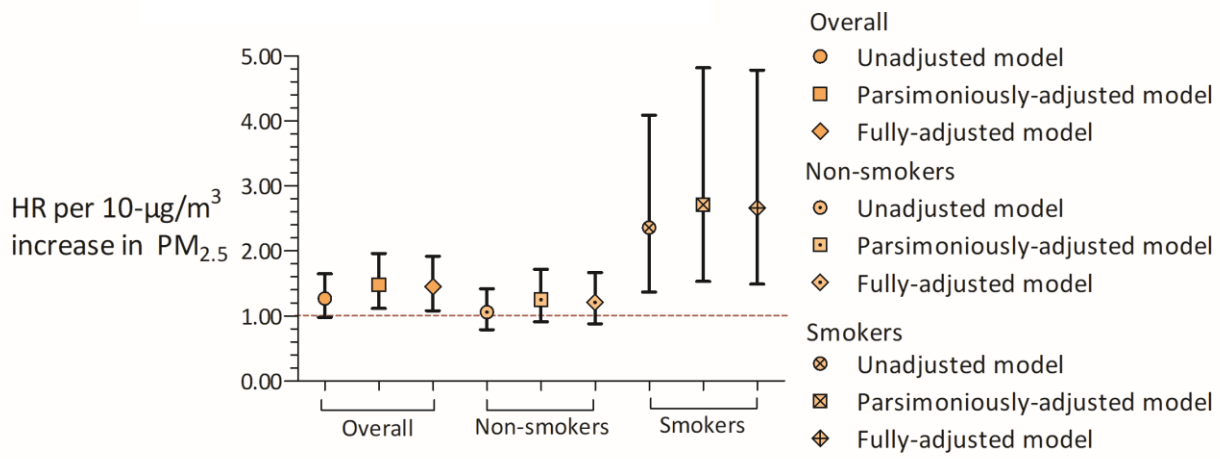
[a]Covariates in the unadjusted model include $PM_{2.5}$ only.

[b]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, passive smoking status, vegetable consumption, history of disease, and occupation.

[c]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Abbreviations: HR, hazard ratio, per 10-μg/m³ increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (kg/m²).

Figure S2.2. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Smoking Status



Abbreviations: HR, hazard ratio, per 10-μg/m³ increase in $PM_{2.5}$ with 95% confidence intervals.

Table S2.7. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Body Mass Index

Variables in three-year average exposure model	Overall		Body Mass Index					
			Normal or underweight		Overweight		Obese	
	(Total cases = 386)		(No. of cases = 226)		(No. of cases = 134)		(No. of cases = 26)	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Unadjusted model[a]	1.27	0.98, 1.65	1.75	1.23, 2.49	0.91	0.59, 1.41	0.34	0.11, 1.08
Parsimoniously-adjusted model[b]	1.45	1.10, 1.92	2.00	1.38, 2.90	1.02	0.64, 1.62	0.25	0.08, 0.83
Fully-adjusted model[c]	1.43	1.07, 1.89	1.93	1.32, 2.82	0.98	0.61, 1.57	0.25	0.08, 0.79

The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%. the fully-adjusted model includes all covariates chosen a priori based on the current literature.

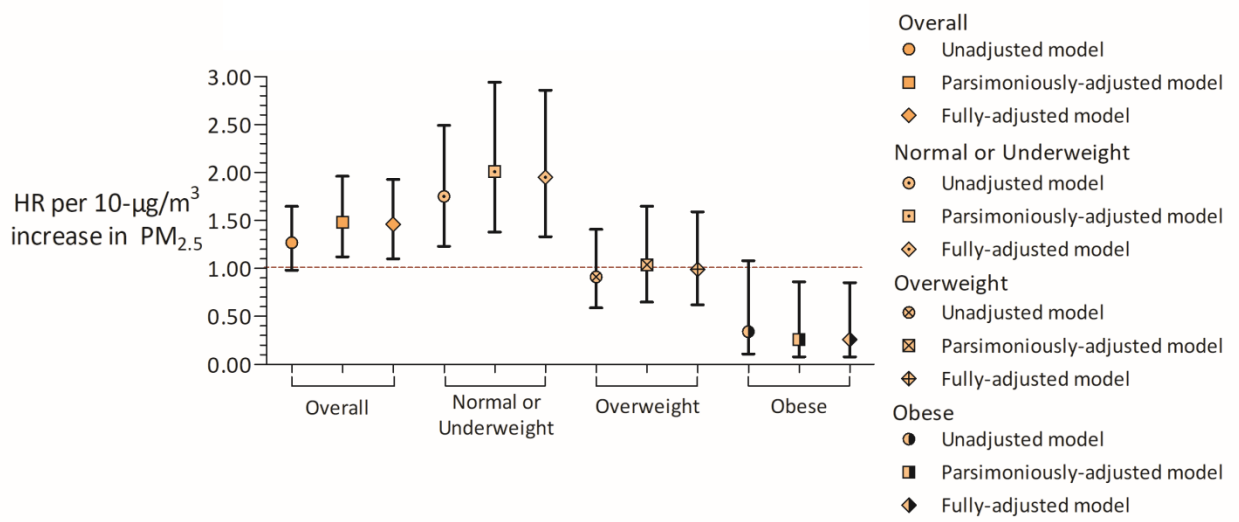
[a]Covariates in the unadjusted model include $PM_{2.5}$ only.

[b]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, passive smoking exposure, vegetable consumption, history of disease, and occupation.

[c]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, exercise, smoking status, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Abbreviations: HR, hazard ratio, per 10-$\mu g/m^3$ increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (kg/m^2), categorized as underweight or normal weight (<24.9), overweight (25.0–29.9), and obese (≥30).

Figure S2.3. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Body Mass Index



Abbreviations: HR, hazard ratio, per 10-µg/m³ increase in $PM_{2.5}$ with 95% confidence intervals.

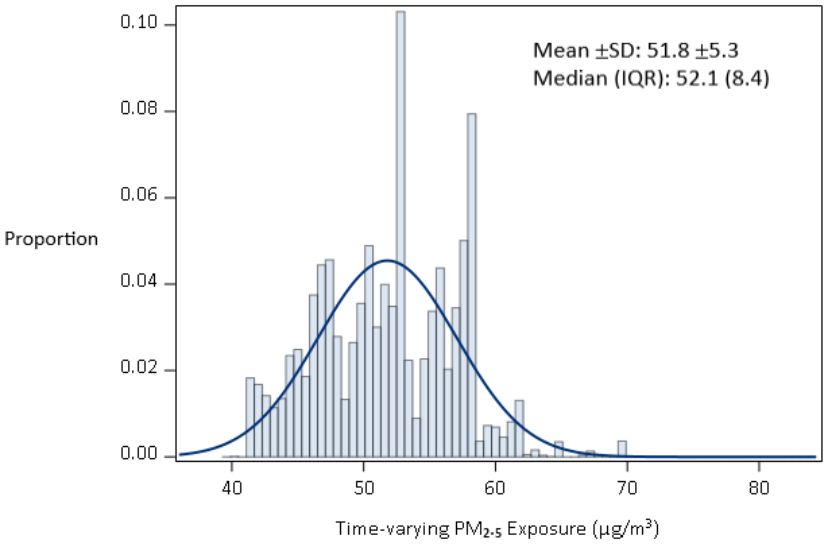
Supplementary Material 3. Yearly Time-Varying $PM_{2.5}$

Table S3.1. Descriptive Statistics for Time-Varying Concentration of Ambient Fine Particulate Matter and Its Constituents in Deqing County

Time-varying concentration (μg/m³)	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
$PM_{2.5}$	51.80	5.27	39.42	47.58	52.08	56.00	84.00	8.42

Abbreviations: SD, standard deviation; IQR, interquartile range; $PM_{2.5}$, ambient fine particulate matter.

Figure S3.1. Histogram of Time-Varying Ambient Fine Particulate Matter Exposure in Deqing County, from 2006 to 2014



Abbreviations: $PM_{2.5}$, ambient fine particulate matter, measured in units of μg/m³; SD, standard deviation; IQR, interquartile range.

Table S3.2. Effect Modification of Time-varying Exposure to Ambient Fine Particulate Matter and Selected Baseline Characteristics on Incidence of Type 2 Diabetes Mellitus

Model	Variables in time-varying exposure model	β	SE	p-value for interaction
1	**Unadjusted model**: $PM_{2.5}$	—	—	—
2	$PM_{2.5}$ + Gender + ($PM_{2.5}$ × Gender)			0.1525
	Male vs. Female	0.0271	0.0189	
3	$PM_{2.5}$ + Age + ($PM_{2.5}$ × Age)			0.9917
	<40 vs. 50–59	0.0008	0.0572	

Model	Variables in time-varying exposure model	β	SE	p-value for interaction
	40–49 vs. 50–59	0.0069	0.0248	
	60–69 vs. 50–59	0.0014	0.0248	
	≥70 vs. 50–59	-0.0109	0.0312	
4	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0039
	Ever smoked vs. Never smoked	0.0622	0.0215	
5	$PM_{2.5}$ + BMI + ($PM_{2.5}$ × BMI)			0.5414
	Overweight vs. Normal or underweight	-0.01622	0.01989	
	Obese vs. Normal or underweight	-0.03379	0.03793	

A p-value for interaction of <0.05 was considered statistically significant and an indicator of potential effect modification. Reference categories were chosen based on category with largest sample size in order to ensure stable results. All covariates presented as potential effect modifiers were selected a priori based on the current literature.
Abbreviations: β*, regression coefficient for* $PM_{2.5}$*; SE, standard error;* $PM_{2.5}$*, ambient fine particulate matter; BMI, body mass index (*kg/m^2*), categorized as normal or underweight (<24.9), overweight (25.0–29.9), and obese (≥30).*

Table S3.3. Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Time-Varying Exposure to Ambient Fine Particulate Matter, Adjusted Independently By Potential Confounders

Model	Variables in time-varying exposure model	HR	95% CI	p-value	β	Percent change
1	**Unadjusted model**: $PM_{2.5}$	1.55	1.19, 2.02	0.0011	0.04376	—
2	$PM_{2.5}$ + Gender	1.54	1.18, 2.01	0.0013	0.04330	1.06
3	$PM_{2.5}$ + Age	1.46	1.12, 1.92	0.0057	0.03806	14.98
4	$PM_{2.5}$ + BMI	1.50	1.15, 1.95	0.0026	0.04037	8.40
5	$PM_{2.5}$ + Exercise	1.54	1.18, 2.01	0.0014	0.04313	1.46
6	$PM_{2.5}$ + Smoking status	1.53	1.18, 2.00	0.0015	0.04274	2.39
7	$PM_{2.5}$ + Passive smoking exposure	1.55	1.19, 2.01	0.0013	0.04350	0.60
8	$PM_{2.5}$ + Vegetable consumption	1.62	1.24, 2.12	0.0004	0.04832	9.44
9	$PM_{2.5}$ + Fruit consumption	1.52	1.16, 1.98	0.0022	0.04168	4.99
10	$PM_{2.5}$ + Family history of diabetes	1.58	1.21, 2.05	0.0007	0.04560	4.04
11	$PM_{2.5}$ + History of disease	1.60	1.23, 2.08	0.0004	0.04690	6.70
12	$PM_{2.5}$ + Education	1.48	1.14, 1.94	0.0038	0.03939	11.09
13	$PM_{2.5}$ + Occupation	1.63	1.25, 2.12	0.0003	0.04876	10.25
14	$PM_{2.5}$ + Family income	1.54	1.19, 2.01	0.0012	0.04343	0.76
15	**Parsimoniously-adjusted model**[a]	1.48	1.13, 1.94	0.0049	—	—

Percent change of regression *coefficient was calculated as |(nested model* β *– fuller model* β*) ÷ fuller model* β*|×100%. A percent change of* regression *coefficient ≥10% was considered strong evidence of confounding of the association between long-term exposure to ambient fine particulate matter and incidence of type 2 diabetes mellitus. All covariates presented as potential confounders were selected a priori based on the current literature. the parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%.*
[a]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$*, age, education, and occupation.*
Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$*; CI, confidence interval;* β*, regression coefficient;* $PM_{2.5}$*, ambient fine particulate matter; BMI, body mass index.*

Table S3.4. Effect Modification of Time-Varying Exposure to Ambient Fine Particulate Matter in Parsimoniously- and Fully-Adjusted Models

Model	Variables in baseline exposure model	β	SE	p-value for interaction
1	**Parsimoniously-adjusted model**[a]	—	—	—
2	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0057
	Ever smoked vs. Never smoked	0.05967	0.02158	
3	**Fully-adjusted model**[b]			
4	$PM_{2.5}$ + Smoking status + ($PM_{2.5}$ × Smoking status)			0.0068
	Ever smoked vs. Never smoked	0.05832	0.02154	

[a]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, education, and occupation.

[b]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Table S3.5. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Time-Varying Exposure to Ambient Fine Particulate Matter, Stratified By Smoking Status

Variables in time-varying exposure model	Overall		Smoking Status			
			Non-smokers		Smokers	
	(Total cases = 386)		(No. of cases = 308)		(No. of cases = 78)	
	HR	95% CI	HR	95% CI	HR	95% CI
Unadjusted model[a]	1.55	1.19, 2.02	1.30	0.97, 1.75	2.65	1.52, 4.61
Parsimoniously-adjusted model[b]	1.48	1.13, 1.94	1.25	0.92, 1.70	2.49	1.42, 4.36
Fully-adjusted model[c]	1.51	1.14, 1.99	1.32	0.96, 1.80	2.42	1.36, 4.30

The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%. the fully-adjusted model includes all covariates chosen a priori based on the current literature.

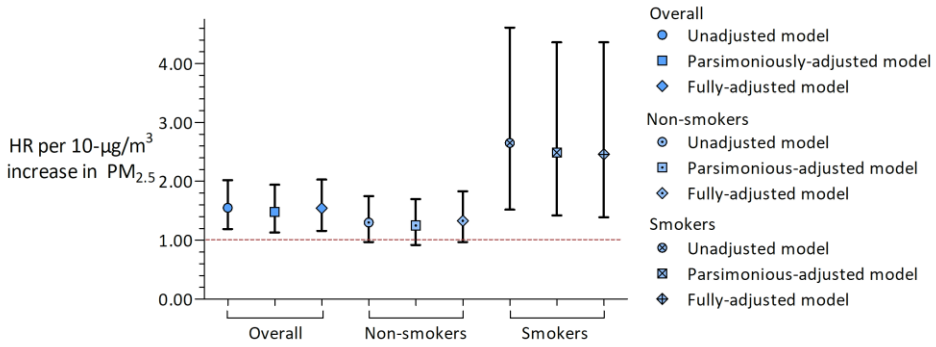
[a]Covariates in the unadjusted model include $PM_{2.5}$ only.

[b]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, education, and occupation.

[c]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, BMI, exercise, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (kg/m^2).

Figure S3.2. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Time-Varying Exposure to Ambient Fine Particulate Matter, Stratified By Smoking Status



Abbreviations: HR, hazard ratio, per 10-µg/m³ increase in $PM_{2.5}$ with 95% confidence intervals.

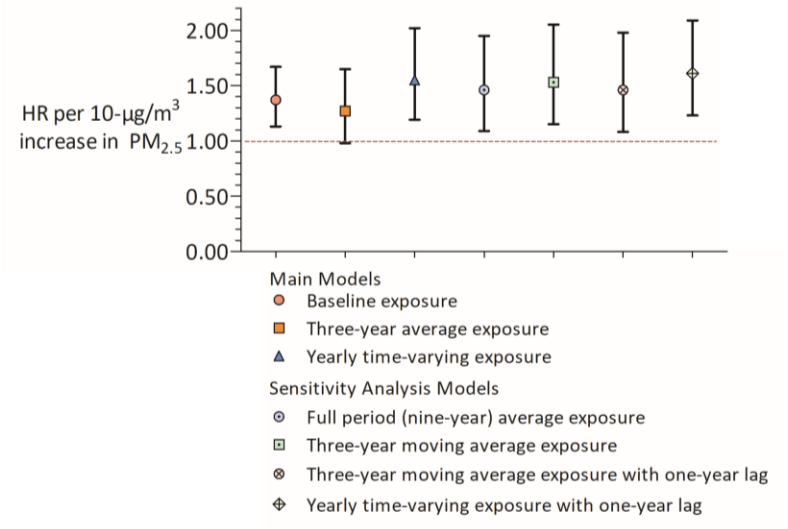
Supplementary Material 4. Sensitivity Analyses

Table S4.1. Sensitivity Analysis for Ambient Fine Particulate Matter Exposure Assignment

Sensitivity analyses	HR	95% CI	p-value	β	SE
Effect of time-fixed vs. time-varying					
Baseline exposure	1.37	1.13, 1.67	0.0014	0.03156	0.00990
Yearly time-varying exposure	1.55	1.19, 2.02	0.0011	0.04376	0.01342
Three-year average exposure	1.27	0.98, 1.65	0.0747	0.02391	0.01342
Three-year moving average exposure	1.53	1.15, 2.05	0.0041	0.04263	0.01486
Effect of single-year vs. average					
Baseline exposure	1.37	1.13, 1.67	0.0014	0.03156	0.00990
Three-year average exposure	1.27	0.98, 1.65	0.0747	0.02391	0.01342
Full period (nine-year) average exposure	1.46	1.09, 1.95	0.0105	0.03789	0.01481
Yearly time-varying exposure	1.55	1.19, 2.02	0.0011	0.04376	0.01342
Three-year moving average exposure	1.53	1.15, 2.05	0.0041	0.04263	0.01486
Effect of lag					
Yearly time-varying exposure	1.55	1.19, 2.02	0.0011	0.04376	0.01342
Yearly time-varying exposure with one-year lag	1.61	1.23, 2.09	0.0005	0.04728	0.01349
Three-year moving average exposure	1.53	1.15, 2.05	0.0041	0.04263	0.01486
Three-year moving average exposure with one-year lag	1.46	1.08, 1.98	0.0129	0.03814	0.01533

All models in the sensitivity analysis are presented as unadjusted models.

Figure S4.1. Sensitivity Analysis for Ambient Fine Particulate Matter Exposure Assignment



Abbreviations: HR, hazard ratio, per 10-µg/m³ increase in $PM_{2.5}$ with 95% confidence intervals; $PM_{2.5}$, ambient fine particulate matter.

Table S4.2. Sensitivity Analyses for Reduced Cohorts using Parsimoniously-Adjusted Models

Sensitivity analyses	Total Cohort	Cases	HR	95% CI	p-value	β	SE
Sensitivity Analysis 1: Excluding cases in first year of follow-up	28,202	309					
Baseline exposure[a]			1.27	1.02, 1.59	0.0362	0.02394	0.01143
Three-year average exposure[b]			1.61	1.18, 2.18	0.0025	0.04734	0.01567
Yearly time-varying exposure[c]			1.46	1.08, 1.96	0.0141	0.03752	0.01528
Sensitivity Analysis 2: Excluding participants contributing ≤2 person-years	24,513	167					
Baseline exposure			1.60	1.20, 2.14	0.0015	0.04712	0.01480
Three-year average exposure			1.68	1.16, 2.53	0.0131	0.05177	0.02087
Yearly time-varying exposure			1.84	1.21, 2.79	0.0047	0.06069	0.02146

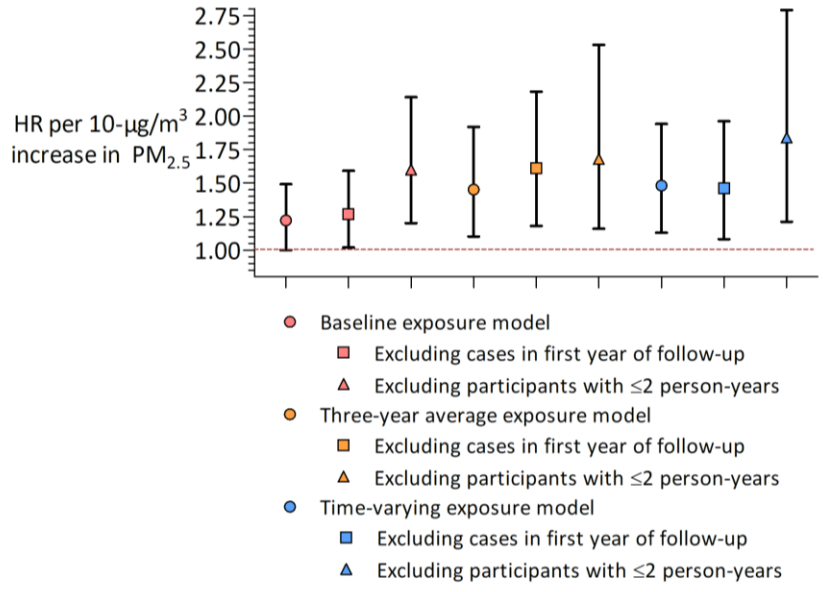
[a]Baseline exposure to $PM_{2.5}$ model, adjusted parsimoniously for age, BMI, and education.

[b]Three-year average exposure to $PM_{2.5}$ model, adjusted parsimoniously for age, passive smoke exposure, vegetable consumption, history of chronic disease(s), and occupation.

[c]Yearly time-varying exposure to $PM_{2.5}$ model, adjusted parsimoniously for age, education, and occupation.

Abbreviations: T2DM, type 2 diabetes mellitus; HR, hazard ratio, per 10-μg/m³ increase in $PM_{2.5}$; CI, confidence interval; β, regression coefficient; SE, standard error; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index.

Figure S4.2. Sensitivity Analyses for Reduced Cohorts using Parsimoniously-Adjusted Models



Abbreviations: HR, hazard ratio, per 10-μg/m³ increase in $PM_{2.5}$ with 95% confidence intervals; $PM_{2.5}$, ambient fine particulate matter.

Table S4.3. Cohort Distribution by BMI defined by International and Chinese Standards, in Full Cohort

Baseline Characteristics	Total cohort		New cases of T2DM		Participants without T2DM at end of follow-up	
	No.	%	No.	%	No.	%
Overall	28,279	100.0	386	100.0	27,893	100.0
Body mass index (kg/m^2), international standard						
Normal or underweight (<24.9)	22,378	79.1	226	58.6	22,152	79.4
Overweight (25.0–29.9)	5,530	19.6	134	34.7	5,396	19.4
Obese (≥30)	371	1.3	26	6.7	345	1.2
Body mass index (kg/m^2), Chinese standard						
Normal or underweight (<23.0)	16,770	59.30	158	40.93	16,612	59.56
Overweight (23.0–27.4)	9,989	35.32	165	42.75	9,824	35.22
Obese (≥27.5)	1,520	5.38	63	16.32	1,457	5.22

Abbreviations: T2DM, type 2 diabetes mellitus.

Table S4.4. Sensitivity Analyses: Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Body Mass Index, Defined by Chinese Standards

Variables in three-year average exposure model	Overall		Body Mass Index, International Standard						Body Mass Index, Chinese Standard					
			Normal or underweight		Overweight		Obese		Normal or underweight		Overweight		Obese	
	(Total cases = 386)		(No. of cases = 226)		(No. of cases = 134)		(No. of cases = 26)		(No. of cases = 158)		(No. of cases = 165)		(No. of cases = 63)	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Unadjusted model[a]	1.27	0.98, 1.65	1.75	1.23, 2.49	0.91	0.59, 1.41	0.34	0.11, 1.08	1.60	1.06, 2.43	1.26	0.84, 1.88	0.59	0.30, 1.14
Parsimoniously-adjusted model[b]	1.45	1.10, 1.92	2.00	1.38, 2.90	1.02	0.64, 1.62	0.25	0.08, 0.83	1.76	1.14, 2.73	1.57	1.02, 2.40	0.61	0.31, 1.22
Fully-adjusted model[c]	1.43	1.07, 1.89	1.93	1.32, 2.82	0.98	0.61, 1.57	0.25	0.08, 0.79	1.69	1.08, 2.64	1.50	0.97, 2.31	0.60	0.30, 1.21

The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%. the fully-adjusted model includes all covariates chosen a priori based on the current literature.

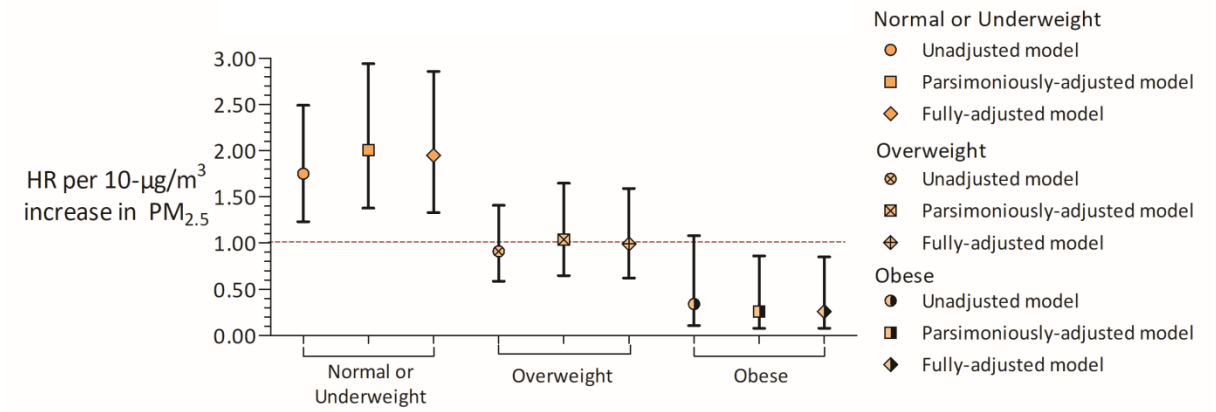
[a]Covariates in the unadjusted model include $PM_{2.5}$ only.

[b]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, passive smoking exposure, vegetable consumption, history of disease, and occupation.

[c]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, exercise, smoking status, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

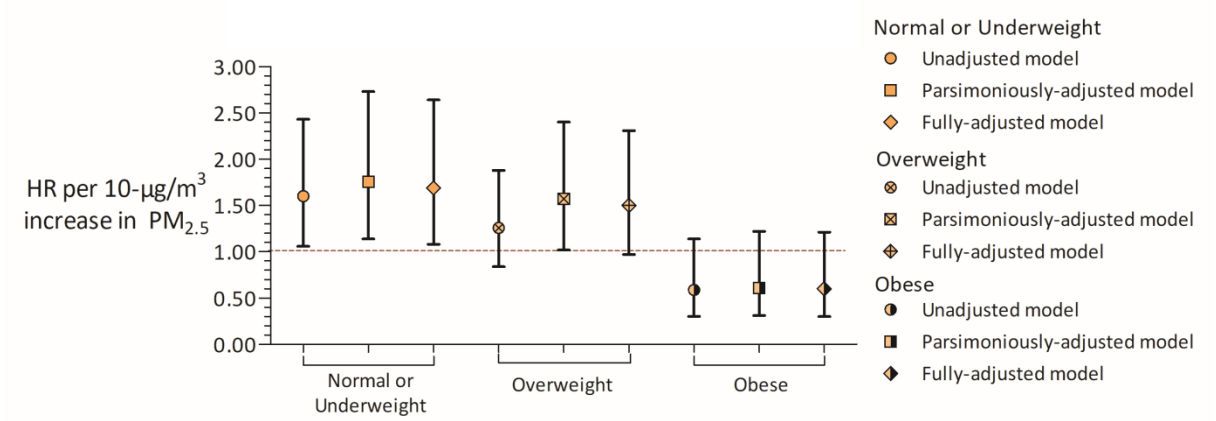
Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BMI, body mass index (kg/m^2), categorized as underweight or normal weight (<24.9), overweight (25.0–29.9), and obese (≥30); Chinese BMI, Chinese standard body mass index (kg/m^2), categorized as normal or underweight (<23.0), overweight (23.0–27.4), and obese (≥27.5)

Figure S4.3. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Body Mass Index, Defined by WHO Standards



Definition: WHO BMI, body mass index (kg/m²), categorized as underweight or normal weight (<24.9), overweight (25.0–29.9), and obese (≥30); HR, hazard ratio, per 10-µg/m³ increase in $PM_{2.5}$ with 95% confidence intervals.

Figure S4.4. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Body Mass Index, Defined by Chinese Standards



Definition: Chinese BMI, Chinese standard body mass index (kg/m²), categorized as normal or underweight (<23.0), overweight (23.0–27.4), and obese (≥27.5); HR, hazard ratio, per 10-µg/m³ increase in $PM_{2.5}$ with 95% confidence intervals.

Table S4.5. Cohort Distribution by Abdominal Obesity defined by International and Chinese Standards, in Subset Cohort

Baseline Characteristics	Total cohort		New cases of T2DM		Participants without T2DM at end of follow-up	
	No.	%	No.	%	No.	%
Subset Cohort, Overall	11,457	100.00	193	100.00	11,264	100.00
Waist circumference (cm), international standard						
Normal (<102 for males; <88 for females)	10,471	91.39	158	81.87	10,313	91.56
Abdominal obesity (≥102 for males, ≥88 for females)	986	8.61	35	18.13	951	8.44
Waist circumference (cm), Chinese standard						
Normal (<90 for males; <80 for females)	7,777	67.88	101	52.33	7,676	68.15
Abdominal obesity (≥90 for males; ≥80 for females)	3,680	32.12	92	47.67	3,588	31.85

Abbreviations: T2DM, type 2 diabetes mellitus.

Definitions: For men, normal waist circumference is <102 or <90 cm, and abdominal obesity is ≥102 or ≥90 cm, defined for the international and Chinese standards, respectively. For females, normal waist circumference is <88 cm or <80 cm, and abdominal obesity ≥88 or ≥80 cm, defined for the international and Chinese standards, respectively

Table S4.6. Sensitivity Analysis: Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Waist Circumference, in a Subset of Full Cohort

Variables in three-year average exposure model	Overall		Waist Circumference, International Standard				Waist Circumference, Chinese Standard			
			Normal		Abdominal Obesity		Normal		Abdominal Obesity	
	(Total cases = 193)		(No. of cases = 158)		(No. of cases = 35)		(No. of cases = 101)		(No. of cases = 92)	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Unadjusted model[a]	1.16	0.79, 1.70	1.36	0.89, 2.09	0.77	0.33, 1.81	1.70	0.98, 2.96	0.88	0.52, 1.50
Parsimoniously-adjusted model[b]	1.35	0.91, 2.01	1.60	1.03, 2.49	0.79	0.32, 1.95	2.02	1.15, 3.54	0.96	0.55, 1.66
Fully-adjusted model[c]	1.29	0.86, 1.93	1.51	0.96, 2.37	0.80	0.31, 2.04	1.90	1.07, 3.38	0.92	0.52, 1.63

The parsimoniously-adjusted model includes all covariates that had a percent change of regression coefficient ≥10%. the fully-adjusted model includes all covariates chosen a priori based on the current literature.

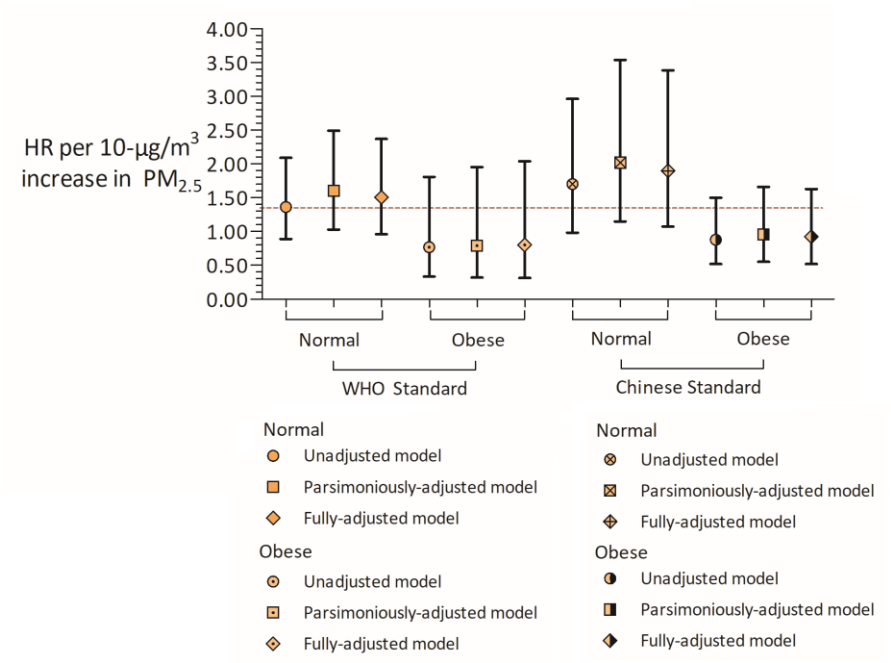
[a]Covariates in the unadjusted model include $PM_{2.5}$ only.

[b]Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, passive smoking exposure, vegetable consumption, history of disease, and occupation.

[c]Covariates in the fully-adjusted model include $PM_{2.5}$, gender, age, exercise, smoking status, passive smoking exposure, vegetable consumption, fruit consumption, family history of diabetes, history of disease, education, occupation, and family income.

Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; WC, International standard waist circumference (cm), categorized as normal (<102 for males; <88 for females) and abdominal obesity (≥102 for males, ≥88 for females); Chinese WC, Chinese standard waist circumference (cm), categorized as normal (<90 for males; <80 for females) and abdominal obesity (≥90 for males; ≥80 for females).

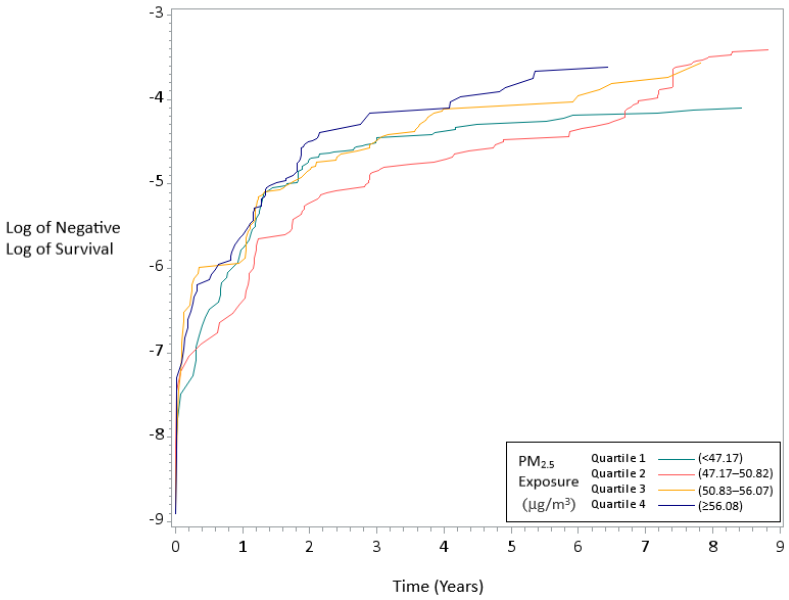
Figure S4.5. Unadjusted and Adjusted Hazard Ratios for Incidence of Type 2 Diabetes Mellitus Associated with Three-Year Average Exposure to Ambient Fine Particulate Matter, Stratified by Abdominal Obesity, Defined by WHO and Chinese Standards



Definitions: For men, normal waist circumference is <102 or <90 cm, and abdominal obesity is ≥102 or ≥90 cm, defined for the international and Chinese standards, respectively. For females, normal waist circumference is <88 cm or <80 cm, and abdominal obesity ≥88 or ≥80 cm, defined for the international and Chinese standards, respectively
Abbreviations: HR, hazard ratio, per 10-µg/m^3 increase in $PM_{2.5}$ with 95% confidence intervals.

Supplementary Material 5. Proportional Hazard Assumption

Figure S5.1. Plot of Log of Negative Log of Survival against Time for Type 2 Diabetes Mellitus Associated with Baseline Exposure Ambient Fine Particulate Matter

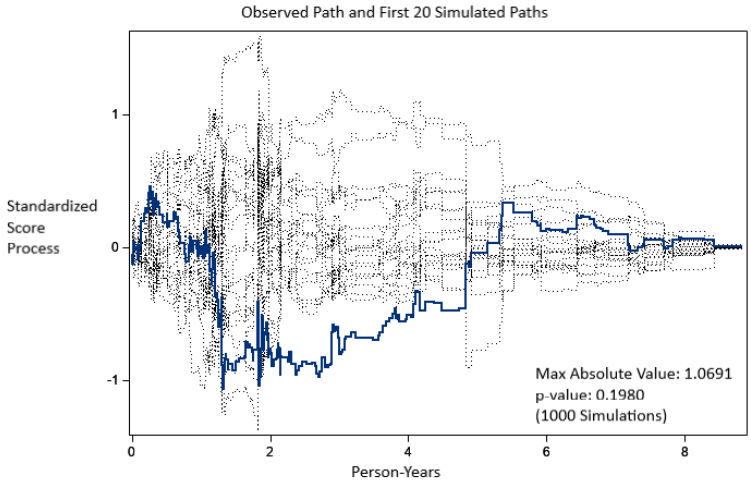


Abbreviations: $PM_{2.5}$, ambient fine particulate matter, categorized into quartile 1 (<47.17 µg/m³), quartile 2 (47.17–50.82 µg/m³), quartile 3 (50.83–56.07 µg/m³), and quartile 4 (≥56.08 µg/m³).

Table S5. Model–Based Method with Time for Type 2 Diabetes Mellitus Associated with Baseline Exposure Ambient Fine Particulate Matter

Variable	β	SE	Chi-square	p-value
($PM_{2.5}$ × time)	0.02475	0.00746	11.0194	0.0009

Figure S5.2. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Ambient Fine Particulate Matter



_____ End of Article 1 _____

CHAPTER 4. Additive Interaction between Long-Term $PM_{2.5}$ Exposure and Smoking Status for Incidence of T2DM (Article 2)

Abstract for Article 2

Background: The association between long-term exposure to ambient fine particulate matter ($PM_{2.5}$) and incidence of type 2 diabetes mellitus (T2DM) has been demonstrated in several cohort studies globally and in our cohort study (**Article 1**). There may be a biological interaction between cigarette smoking and $PM_{2.5}$ on the risk of T2DM.

Objective: The purpose of this article is to assess whether smoking status had noticeable influence on the association between long-term exposure to $PM_{2.5}$ and incidence of T2DM as additive effect modifiers of this association in a rural cohort in China. The follow-up purpose of this article is to assess whether effect modification by smoking status differed when considering gender.

Methods: The same cohort, taken from Deqing County, Zhejiang Province, China, of residents from rural townships and villages aged 18–64 years old and free of T2DM at baseline, was used in the present article. $PM_{2.5}$ exposure was dichotomized using the median as the cut-off point. To assess whether smoking status interacted with $PM_{2.5}$ and T2DM on an additive scale, estimates of relative excess risk due to interaction (RERI), attributable proportion due to interaction (AP), and the synergy index (SI) were calculated and presented with 95% confidence intervals (CI). A p-value of 0.05 was chosen as the level of statistical significance.

Results: There was no statistically significant effect between smoking status and $PM_{2.5}$ exposure on incidence of T2DM in the overall cohort or according to gender. Results remained statistically insignificant regardless of whether the model was unadjusted, parsimoniously-adjusted, or fully-adjusted for potential confounders.

Conclusions: Our study showed no significant additive interaction between cigarette smoking and $PM_{2.5}$ exposure in relation to the incidence of T2DM.

Contents

List of Tables

List of Abbreviations and Acronyms

AP	Attributable proportion due to interaction
CI	Confidence interval
FPG	Fasting plasma glucose
HR	Hazard ratio
IARC	International Agency for Research on Cancer
$PM_{2.5}$	Fine particulate matter
RERI	Relative excess risk due to interaction
SI	Synergy index
T2DM	Type 2 diabetes mellitus

4.1 Introduction

Type 2 diabetes mellitus (T2DM) is a complex disease that, when managed improperly, can lead to multiple acute and chronic negative health impacts.[1] It is estimated that by 2035, 592 million adults worldwide will have T2DM, with sharp increase in low- and middle-income countries such as China.[2] Many factors for T2DM, including family history of T2DM, obesity, and sedentary behaviour, have been well known to increase the risk.[3] There has been growing interest in exploring the relationship between long-term exposure to ambient fine particulate matter ($PM_{2.5}$) and the incidence of T2DM. The association has been demonstrated in previous cohort studies globally and has additionally been examined in our previous article (**Article 1**). In Article 1, long-term exposure to $PM_{2.5}$ was associated with an increased risk of T2DM incidence in a rural cohort in China. It necessarily filled the gap in measuring the association in developing, non-urban populations.

Previous research has provided evidence that there is an interaction between cigarette smoking and $PM_{2.5}$ on the risk of T2DM.[4,5] In the present study, it was hypothesized that inhaling cigarette smoke and breathing $PM_{2.5}$ work synergistically via a biologic interaction for their effects on the development of T2DM. There are possible biological mechanisms underlying how smoking may interact with $PM_{2.5}$ and other inhaled exposures.[5] For example, at the exposure-dose level, smokers and non-smokers may exhibit differing patterns of lung deposition and clearance that may additionally affect deposition and clearance of $PM_{2.5}$. Smokers compared to non-smokers may also exhibit different patterns of activity and ventilation rates that may increase the dose of $PM_{2.5}$ inhaled. Furthermore, at the molecular-level, certain toxic particles in the $PM_{2.5}$ mass mixture may work in the same steps or similar ways as the carcinogens in tobacco smoke.[5]

While the interaction between smoking and exposure to $PM_{2.5}$ has been previously examined on a multiplicative scale,[6] the additive scale has been suggested as the better choice when an interaction is hypothesized to be biologic in nature.[7–9]

In this study, whether the association between long-term exposure to $PM_{2.5}$ and incidence of T2DM was modified by smoking status on an additive scale was examined in a rural Chinese cohort. Effect modification was further examined stratified by gender. Gender-specific analysis for smoking status has been previously implemented in other published research.[10] Gender distributions of smoking status are uniquely integrated in the culture of Chinese society, as smoking is common in males but not in females.[11]

4.2 Methods

Study Design and Population

The cohort was based on secondary data from the Rural Deqing Cohort Study; participants were recruited during 2006–2014 for the original purpose of creating a T2DM rural risk assessment tool.[12] Adult residents, aged 18 to 64 years old, were randomly cluster sampled from eight rural townships, included 65 villages in Deqing County, Zhejiang Province, China.[13] Recruited participants were given a baseline questionnaire to collect details on demographics and potential covariate data such as smoking status, and were given a physical examination, in which a blood sample was drawn, to identify cases of T2DM over the study period. Participants included in the analytical dataset were all free of T2DM at baseline. Further details about the cohort, data collection, and study design are presented in the previous article (**Article 1**).

Exposure Assessment

Satellite-derived estimates of ambient $PM_{2.5}$ concentrations were used to approximate ground-level $PM_{2.5}$ using the V4.CH.02 product of the Dalhousie University Atmospheric Composition Analysis Group.[14] Ground-based observations from the recently expanded monitoring network over mainland China incorporated monthly geographically weight regressions at a 0.01° × 0.01° resolution from 2014 to 2016. The resultant annual average $PM_{2.5}$ concentrations were consistent with out of sample, cross-validation observations ($R^2 = 0.78$). $PM_{2.5}$ concentrations prior to this time period are based on the relative changes of Dalhousie University Atmospheric Composition Analysis Group's long-term product.[14–16]

$PM_{2.5}$ exposure was assigned using each participant's home residence on a 1 × 1 km resolution and assigned as the concentration in the baseline year as a surrogate for total exposure. Participants with missing information on exposure were excluded from the analytical dataset.

Outcome and Covariates

Incident cases of T2DM were identified by the results of a fasting plasma glucose (FPG) test, self-reported diagnosis of T2DM, or use of anti-diabetic medication. FPG test was determined from a blood sample collected during follow-up physical examinations after an overnight fast of at least 8 hours. FPG of 7.0 mmol/L or higher was the cut-off for defining cases of T2DM. Incident cases of T2DM were ascertained from the Deqing electronic health records until December 31, 2018.

Smoking status was identified from responses in the baseline questionnaire and dichotomized as 'non-smoker' or 'smoker' in the analytical dataset. Participants were considered smokers if they smoked at least one cigarette per day for greater than six months. Participants with missing information on

smoking status were excluded from the analytical dataset. For further details on the covariates, please see the previous article (**Article 1**).

4.3 Statistical Analysis

Main Analysis

Cox proportional hazards models were used to determine the association between long-term exposure to $PM_{2.5}$ and the incidence of T2DM and hazard ratios (HR) with 95% confidence intervals (CI) were calculated. Effect modification of smoking status on the relationship between long-term exposure to $PM_{2.5}$ and incidence of T2DM was conducted on both multiplicative and additive scales. Effect modification of smoking status on the multiplicative scale was conducted by including interaction terms and generating a p-value for the interaction, for which a p-value for the interaction less than 0.05 was considered significant. The additive interaction between smoking status and $PM_{2.5}$ was evaluated using three indexes: relative excess risk due to interaction (RERI), attribution proportion of interaction (AP), and the synergy index (SI).[17,18] RERI (Equation 1) is interpreted as the increased hazard due to the additive interaction as a proportion of the hazard of both risk factors at level 0.[17]

Equation 1: $$RERI = HR_{11} - HR_{10} - HR_{01} + 1$$

AP (Equation 2) is interpreted as the total hazard due to the additive interaction itself.[17,18]

Equation 2: $$AP = \frac{RERI}{HR_{11}}$$

SI (Equation 3) is interpreted as the ratio of increase in hazard due to both risk factors to the sum of increases due to each risk factor separately.[17,18]

Equation 3: $$SI = \frac{HR_{11}-1}{(HR_{10}-1)+(HR_{01}-1)}$$

In the absence of an interaction effect, RERI is equal to 0, AP is equal to 0, and SI is equal to 1.[19]

A pre-established SAS macro by Li & Chambless was used to calculate RERI, AP, and SI in Cox proportional hazards models.[17] To include variables in this SAS macro, both variables had to be dichotomous. Smoking status was already a dichotomous variable, comparing 'smokers' to 'non-smokers'. The continuous $PM_{2.5}$ variable was dichotomized using the median value as the cut-off. In the additive interaction model, new dummy variables were created to represent the independent and joint effect of $PM_{2.5}$ exposure and smoking status. Specifically, the newly created variable of joint exposure were: non-smoker with low $PM_{2.5}$ exposure (the reference category), smoker with low $PM_{2.5}$ exposure (i.e. the effect due to smoking), non-smoker with high $PM_{2.5}$ exposure (i.e. the effect due to exposure to $PM_{2.5}$), and smoker with high $PM_{2.5}$ exposure (i.e. joint effect due to exposure to $PM_{2.5}$ and smoking). The variable representing smoker with high $PM_{2.5}$ exposure represents the biologic interaction of interest. Effect modification of smoking status was further assessed stratified by gender. Gender-specific analysis for smoking status has been previously implemented in other published research.[10]

All statistical analysis were completed in SAS 9.4.

Sensitivity Analysis

In the sensitivity analysis, $PM_{2.5}$ was dichotomized using the 1st and 3rd quartiles of $PM_{2.5}$ exposure as cut-offs instead of the median, as was done previously in other published research.[4] Results were expected to remain suggestive of an additive interaction if smoking truly exhibited a biologic interaction with $PM_{2.5}$ exposure.

4.4 Results

Description Statistics

There were a total of 28,279 participants in the analytical dataset, with 386 identified cases of T2DM at the end of follow-up (**Table 1**). Detailed descriptions and a table of the baseline characteristics of the cohort can be found in the previous article (**Article 1**). There were a total of 12,513 males and 15,766 females, with 144 and 242 male and female cases of T2DM identified at the end of follow-up (**Table 1**). The median concentration of baseline exposure to $PM_{2.5}$ was 50.83 μg/m^3 and was used as the cut-off for defining low (lower than 50.83 μg/m^3) vs. high (50.83 μg/m^3 and higher) exposure in this analysis. In the overall cohort, more adult participants identified as non-smokers (77.3%) than smokers (22.7%). Expectantly, the distribution of smoking status differed drastically by gender. In males, approximately half were smokers (48.8%). In females, only 2.0% were smokers, with the overwhelming majority identifying as non-smokers (98.0%).

Table 1. Baseline Smoking Status and Ambient Fine Particulate Matter Exposure, Overall and According to Gender

Baseline Characteristics	Overall				Male				Female			
	Total cohort		Cases of T2DM		All males		Cases of T2DM		All females		Cases of T2DM	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Overall	28,279	100.0	386	100.0	12,513	100.0	144	100.0	15,766	100.0	242	100.0
Smoking status												
Non-smokers	21,857	77.3	308	79.8	6,404	51.2	71	49.3	15,453	98.0	237	97.9
Smokers	6,422	22.7	78	20.2	6,109	48.8	73	50.7	313	2.0	5	2.1
$PM_{2.5}$ exposure												
Low (<50.83 μg/m^3)	14,084	49.8	213	55.2	6,266	50.1	81	56.3	7,818	49.6	132	54.6
High (≥50.83 μg/m^3)	14,195	50.2	173	44.8	6,247	49.9	63	43.7	7,948	50.4	110	45.4

The median concentration of baseline exposure to $PM_{2.5}$ (50.83 μg/m^3) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the main analysis.

All values of $PM_{2.5}$ are based on baseline exposure models unless otherwise stated.

Abbreviations: T2DM, type 2 diabetes mellitus; $PM_{2.5}$, ambient fine particulate matter.

The incidence rates of T2DM are presented in **Table 2**, according to smoking status, $PM_{2.5}$ exposure, and gender. Overall and across genders, high $PM_{2.5}$ exposure exhibited a higher incidence compared to participants exposed to low $PM_{2.5}$. Smokers, on the other hand, exhibited a lower incidence per 1000 person-years compared to non-smokers.

Table 2. Incidence Rates of Type 2 Diabetes Mellitus by Smoking Status and Ambient Fine Particulate Matter Exposure, Overall and According to Gender

Baseline Characteristics	Overall				Males				Females			
	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI
Overall	386	118,117.3	3.27	2.96, 3.61	144	52,785.9	2.73	2.32, 3.21	242	65,331.5	3.70	3.27, 4.20
Smoking status												
Non-smokers	308	88,459.4	3.48	3.11, 3.89	71	24,617.2	2.88	2.29, 3.64	237	63,842.2	3.71	3.27, 4.22
Smokers	78	29,657.9	2.63	2.11, 3.28	73	28,168.6	2.59	2.06, 3.26	5	1,489.3	3.36	1.40, 8.08
$PM_{2.5}$ exposure												
Low (<50.83 μg/m^3)	213	74,825.8	2.85	2.49, 3.26	81	33,401.9	2.43	1.95, 3.02	132	41,423.9	3.19	2.69, 3.78
High (≥50.83 μg/m^3)	173	43,291.5	4.00	3.44, 4.64	63	19,383.9	3.25	2.54, 4.16	110	23,907.5	4.60	3.82, 5.55

The median concentration of baseline exposure to $PM_{2.5}$ (50.83 μg/m^3) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the main analysis.

All values of $PM_{2.5}$ are based on baseline exposure models unless otherwise stated.

Abbreviations: T2DM, type 2 diabetes mellitus; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter.

Assessment of Multiplicative Interaction

Results of the assessment of a multiplicative interaction between smoking status and dichotomized $PM_{2.5}$ exposure on incidence of T2DM is presented in **Table 3**. The data showed no significant multiplicative interaction between smoking status and $PM_{2.5}$ exposure on the incidence of T2DM overall (p-value = 0.1800) or for females (p-value = 0.9817) when $PM_{2.5}$ exposure was dichotomized with a median concentration cut-off in the unadjusted model. The unadjusted HR for T2DM associated with high $PM_{2.5}$ exposure (≥50.83 μg/m^3) vs. low exposure (<50.83 μg/m^3) in the overall cohort was 1.21 (95% CI: 0.96, 1.52) for non-smokers and 1.69 (95% CI: 1.08, 2.65) for smokers. Given the small sample size per category amongst females (i.e. there are only a total of 5 cases of T2DM amongst female smokers), model results in females are presented but not interpreted. Although no significant multiplicative interaction was identified for the overall cohort and for females, there was a significant multiplicative interaction between smoking status and $PM_{2.5}$ exposure in males (p-value = 0.0484); the unadjusted HR for T2DM associated with high $PM_{2.5}$ in males was 0.89 (95%: 0.54, 1.44) for non-smokers and 1.73 (95% CI: 1.08, 2.76) for smokers.

Table 3. Multiplicative Effect Modification of Smoking Status and Baseline Exposure to Ambient Fine Particulate Matter on Incidence of Type 2 Diabetes Mellitus, Overall and According to Gender

	Cases of T2DM	Unadjusted model			Parsimoniously-adjusted model			Fully-adjusted model		
		HR	95% CI	p-value for interaction	HR	95% CI	p-value for interaction	HR	95% CI	p-value for interaction
Overall	**386**			0.1830			0.2326			0.2442
Non-smokers	308									
Low $PM_{2.5}$ (<50.83 μg/m^3)	173	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥50.83 μg/m^3)	135	1.21	0.96, 1.52		1.06	0.84, 1.34		1.09	0.85, 1.38	
Smokers	78									
Low $PM_{2.5}$ (<50.83 μg/m^3)	40	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥50.83 μg/m^3)	38	1.69	1.08, 2.65		1.43	0.91, 2.25		1.46	0.93, 2.30	
Male	**144**			0.0484			0.0696			0.0932
Non-smokers	71									
Low $PM_{2.5}$ (<50.83 μg/m^3)	44	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥50.83 μg/m^3)	27	0.89	0.54, 1.44		0.79	0.48, 1.29		0.82	0.50, 1.35	
Smokers	73									
Low $PM_{2.5}$ (<50.83 μg/m^3)	37	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥50.83 μg/m^3)	36	1.73	1.08, 2.76		1.46	0.91, 2.35		1.45	0.90, 2.33	
Female[a]	**242**			0.2140			0.9841			0.9690
Non-smokers	237									
Low $PM_{2.5}$ (<50.83 μg/m^3)	129	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥50.83 μg/m^3)	108	1.33	1.02, 1.73		1.15	0.88, 1.51		1.19	0.90, 1.57	
Smokers	5									
Low $PM_{2.5}$ (<50.83 μg/m^3)	3	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥50.83 μg/m^3)	2	1.35	0.23, 8.11		1.17	0.20, 7.01		1.23	0.21, 7.40	

The median concentration of baseline exposure to $PM_{2.5}$ (50.83 μg/m^3) was used as cut-off for low vs. high. A p-value for interaction of <0.05 was considered statistically significant and evidence of multiplicative effect modification.

[a]Given the small sample size per category amongst females, model results in females are presented but not interpreted.

Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter.

In the parsimoniously-adjusted model controlling for age, BMI, and education, the multiplicative interaction between smoking status and $PM_{2.5}$ exposure on incidence of T2DM was insignificant for the cohort overall (p-value = 0.2326) and for females (p-value = 0.9841), and borderline significant for males (p-value = 0.0696). The parsimoniously-adjusted HR for T2DM associated with high $PM_{2.5}$ exposure in males was 1.46 (95% CI: 0.91, 2.35) in smokers and 0.79 (95% CI: 0.48, 1.29) in non-smokers.

In the fully-adjusted model controlling for all gender (in overall analysis only), age, BMI, exercise, passive smoking exposure, vegetable and fruit consumption, family history of T2DM, history of disease(s), education, occupation, and family income, the multiplicative interaction was also insignificant for the cohort overall (p-value = 0.2442), for females (p-value = 0.9690), and for males (p-value = 0.0932). In general, disregarding confidence, the results from varying degrees of model adjustment suggest similar patterns; whereby smokers with high $PM_{2.5}$ exposure have the greatest risk, regardless of gender. Nonetheless, the lack of statistical significance and CIs containing the null hypothesis impede any substantial conclusions.

Assessment for Joint Effect

Evidence of a joint effect is presented in **Table 4**. The data suggested that there was no statistically significant effect between smoking status and $PM_{2.5}$ exposure on the incidence of T2DM overall or according to gender, regardless of degree of model adjustment. In the parsimoniously-adjusted model, the joint effect of smoking and high $PM_{2.5}$ exposure demonstrated no association (HR = 1.03, 95% CI: 0.72, 1.46) in the overall cohort and modest but still insignificant associations (HR = 1.14, 95% CI: 0.73, 1.78) in males. Due to low sample size per category, the results for females are not described. In the fully-adjusted model, the joint effect of smoking and high $PM_{2.5}$ exposure resulted in a

HR of 1.39 (95% CI: 0.93, 2.08) compared to the reference group of non-smokers with low $PM_{2.5}$ exposure. In males, the HR for joint effect was 1.23 (95% CI: 0.78, 1.94).

Table 4. Joint Effect of Smoking Status and Baseline Exposure to Ambient Fine Particulate Matter on Incidence of Type 2 Diabetes Mellitus, Overall and According to Gender

	Cases of T2DM	Unadjusted model		Parsimoniously-adjusted model		Fully-adjusted model	
		HR	95% CI	HR	95% CI	HR	95% CI
Overall	**386**						
Non-smokers and low $PM_{2.5}$	173	1.00	Ref	1.00	Ref	1.00	Ref
Independent effect of smoking	40	0.66	0.47, 0.94	0.72	0.51, 1.01	0.89	0.60, 1.32
Independent effect of high $PM_{2.5}$	135	1.21	0.96, 1.52	1.06	0.84, 1.34	1.16	0.91, 1.47
Joint effect of smoking and high $PM_{2.5}$	38	1.12	0.79, 1.60	1.03	0.72, 1.46	1.39	0.93, 2.08
Male	**144**						
Non-smokers and low $PM_{2.5}$	44	1.00	Ref	1.00	Ref	1.00	Ref
Independent effect of smoking	37	0.70	0.45, 1.09	0.78	0.50, 1.21	0.81	0.52, 1.27
Independent effect of high $PM_{2.5}$	27	0.89	0.54, 1.44	0.79	0.48, 1.29	0.87	0.53, 1.43
Joint effect of smoking and high $PM_{2.5}$	36	1.21	0.78, 1.89	1.14	0.73, 1.78	1.23	0.78, 1.94
Female[a]	**242**						
Non-smokers and low $PM_{2.5}$	129	1.00	Ref	1.00	Ref	1.00	Ref
Independent effect of smoking	3	0.93	0.30, 2.91	0.98	0.31, 3.08	1.03	0.33, 3.36
Independent effect of high $PM_{2.5}$	108	1.33	1.02, 1.73	1.15	0.88, 1.51	1.27	0.97, 1.67
Joint effect of smoking and high $PM_{2.5}$	2	1.26	0.31, 0.51	1.14	0.28, 4.63	1.44	0.36, 5.83

The median concentration of baseline exposure to $PM_{2.5}$ (50.83 µg/m^3) was used as cut-off for low vs. high. A p-value for interaction of <0.05 was considered statistically significant and evidence of multiplicative effect modification.
[a]Given the small sample size per category amongst females, model results in females are presented but not interpreted.
Abbreviations: HR, hazard ratio, per 10-µg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter.

The assessment of an additive interaction using RERI, AP, and SI statistics is presented in **Table 5** for unadjusted, parsimoniously-adjusted, and fully-adjusted models. In the absence of an interaction effect (i.e. the null hypothesis is true), RERI is equal to 0, AP is equal to 0, and SI is equal to 1.[19] In the parsimoniously-adjusted model, the RERI was 0.003 (95% CI: -0.26, 0.27), the AP was 0.004 (95% CI: -0.45, 0.46) and the SI was 0.99 (95% CI: 0.29, 3.29). In males specifically, the RERI was comparable. In the fully-adjusted model, the RERI was 0.39 (95% CI: -0.32, 1.09), the AP was 0.27 (95% CI: -0.10, 0.64), and the SI was 9.23 (95% CI: 0.002, 3.74×10^4) for the overall cohort, indicating no significant additive interaction between smoking status and $PM_{2.5}$ exposure on the incidence of T2DM. In males, the RERI

was 0.18 (95% CI: -0.22, 0.59), the AP was 0.21 (95% CI: -0.35, 0.77), and the SI was 0.43 (95% CI: 0.003, 60.67).

Table 5. Measure of Additivity for Smoking Status and Baseline Exposure to Ambient Fine Particulate Matter on Incidence of Type 2 Diabetes Mellitus, Overall and According to Gender

Measure of Additivity	Overall		Male		Female[a]	
	Point estimate	95% CI	Point estimate	95% CI	Point estimate	95% CI
Unadjusted model						
RERI	0.028	-0.30, 0.35	0.16	-0.16, 0.48	0.29	-2.00, 2.58
AP	0.031	-0.40, 0.47	0.22	-0.35, 0.78	0.19	-1.02, 1.39
SI	0.79	0.024, 26.29	0.60	0.090, 4.04	2.15	0.04, 126.50
Parsimoniously-adjusted model						
RERI	0.003	-0.26, 0.27	0.13	-0.15, 0.42	0.16	-1.68, 2.00
AP	0.004	-0.45, 0.46	0.19	-0.41, 0.78	0.12	-1.18, 1.43
SI	0.99	0.29, 3.29	0.69	0.19, 2.58	2.25	0.001, 3,656.37
Fully-adjusted model						
RERI	0.39	-0.32, 1.09	0.18	-0.22, 0.59	0.59	-2.28, 3.45
AP	0.27	-0.10, 0.64	0.21	-0.35, 0.77	0.31	-0.71, 1.33
SI	9.23	0.002, 37,390.64	0.43	0.003, 60.67	2.90	0.11, 78.19

The median concentration of baseline exposure to $PM_{2.5}$ (50.83 µg/m^3) was used as cut-off for low vs. high.

[a]Given the small sample size per category amongst females, model results in females are presented but not interpreted.

Abbreviations: CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; RERI, relative excess risk for interaction; AP, attributable proportion of interaction; SI, synergy index.

Sensitivity analysis was performed using alternative cut-off values for $PM_{2.5}$. In the sensitivity analysis, the 1st quartile of $PM_{2.5}$ concentration (47.10 µg/m^3) was used as the cut-off a new definition of low and high $PM_{2.5}$ exposure in **Supplementary Material 1**. Results aligned with the findings in the main analysis.

4.5 Discussion

Summary

There was no clear evidence of effect modification on a multiplicative scale by smoking status on the relationship between T2DM and $PM_{2.5}$ exposure when exposure was dichotomized. There was also no evidence supporting hypothetical synergism between smoking status and $PM_{2.5}$ exposure based on results on an additive scale. Limitations include the requirement to dichotomize $PM_{2.5}$ exposure to assess an additive interaction, as well as limited power to detect a change due to inadequate sample size. Studies of a larger scale and more complex statistically methods are required to further elucidate on the potential biologic effect of smoking and $PM_{2.5}$ exposure on the risk of developing T2DM.

Comparison to Other Research

In their study on cardiovascular mortality and long-term exposure to $PM_{2.5}$, Pope et al. (2004) concluded that "Mechanisms by which cigarette smoke and air particulate exposure operate for these cardiovascular causes of death may be complementary and seem to be at least additive if not synergistic".[20] Despite such words, the joint effect between smoking and $PM_{2.5}$ has not been frequently assessed. A study on lung cancer mortality by Turner et al. (2014) assessed whether different cut-offs of

$PM_{2.5}$ exposure, such as at the 50th percentile or 75th vs. 25th percentile, affected results of the combined effect, and found that the use of 75th and 25th percentiles as cut-offs led to increased evidence of a joint effect in men. Authors described evidence of an additive interaction between smoking and high $PM_{2.5}$ exposure on lung cancer mortality.[4] In a study on adult pulmonary function, associations between $PM_{2.5}$ and pulmonary function were significantly greater among smokers vs. non-smokers, suggesting a synergistic effect of air pollution and smoking on adult pulmonary function.[21]

To our knowledge, there exists only one other study that assessed additive interactions in relation to long-term $PM_{2.5}$ exposure and T2DM.[10] In their nationwide Chinese cohort, Liang et al. (2019) examined interactions on multiplicative and additive scales for age (≥65 vs. <65 years old), gender (female vs. male), urbanity (rural vs. urban), BMI (≥25 vs <25), hypertension (yes vs. no), and smoking (smoker vs. non-smoker). Similar to the present study, they used the p-value for interaction for the HR with a cross-product term for multiplicative interaction, and a p-value for the RERI for the additive interaction. For smoking, authors found that the RERI was -0.11 (95% CI: -0.16, -0.06), in contrast to our cohort's RERI of 0.028 (95%: -0.30, 0.35).

Biological Plausibility

It has been hypothesized that the underlying mechanism by which smoking and $PM_{2.5}$ lead to adverse health effects, such as inflammatory responses, may be similar and therefore work in conjunction with one another. It has additionally been hypothesized that smoking leads to decreased clearance and increased deposition and retention of $PM_{2.5}$, thereby enhancing the adverse effects of $PM_{2.5}$ on health.[22] Smoking may negatively affect the respiratory defense mechanisms needed to clear inhaled $PM_{2.5}$ as well as alter the response to deposited material.[21] In the International Agency for

Research on Cancer (IARC) Scientific Publication on Air Pollution and Cancer, three potential biological mechanisms for the effect modification of $PM_{2.5}$ and smoking were posed that are applicable to the current study. Firstly, the physical activity level and degree of ventilation may differ in smokers compared to those who do not smoke. Secondly, the patterns of lung disposition and clearance are different among smokers compared to non-smokers, and may lead to differences in patterns of lung disposition and clearance in $PM_{2.5}$ as a result. Thirdly, the morphometry of target cells differ in smokers expectedly due to many years of cigarette smoking, and this may affect how these cells react to $PM_{2.5}$ exposure.[5]

Considerations

There are some important limitations of our study that need to be acknowledged. Compared to other nationwide cohort studies which suggest roughly two-thirds of men in China were smokers,[11] the smoking prevalence in males in our cohort appeared to be lower. This suggests that the prevalence of smoking may be lower in our study population compared to the national average. Cohort studies on rural Chinese populations suggest that the smoking prevalence is higher in urban than in rural males.[23,24] The drastically low prevalence of smoking in females was also comparable to the national average. However, in the study population, this led to small sample sizes when stratified by gender in the female category, impeding any confident conclusions from being made.

Importantly, our study may have suffered from limited power and limited precision to detect an effect. The study size was sufficient in addressing the effect in a single exposure, but through interaction analyses, the study cohort was required to be divided into many smaller groups to create contrasting subgroups.[9] As a result, interaction analyses suffered from having less precision than was in the main study analyses (**Article 1**). Furthermore, simplifying the assumptions surrounding the interaction are

harder to justify when risk factors are continuous in true nature,[9] such is the case with $PM_{2.5}$ exposure. The RERI, AP, and S results depend on the chosen reference category.[19] In particular, given that $PM_{2.5}$ exposure is continuous, this means that the RERI, AP, and S are very sensitive to choice of cut-off for high and low exposure. Such a cut-off is not universally known. Perhaps there exists a medically important cut-off that would indicate at what $PM_{2.5}$ exposure level does risk of T2DM increase to a medically important degree; such a number may be a useful choice in cut-off for interaction analysis. But currently, none exist for $PM_{2.5}$ as it relates to T2DM.

An oversimplification of statistical methodology may impede accuracy of results. Particularly, methods of representing a biologic interaction, namely RERI, AP, and S, come with their fair share of limitations. For instance, RERI and AP are not straightforward to interpret when covariates are included into the statistical models for adjusting as confounders. Specifically, the interaction parameter in the model is unvarying, but they actually vary across different covariate strata.[19] Finally, there is some debate on whether an additive scale truly represents a biologic interaction.[9,18]

In addition to the limitations regarding assessment of an additive interaction, there are some important limitations regarding exposure assessment that must also be considered. Quantification of long-term exposure is largely dependent on the spatial variability amongst cohort participants to allow for a noticeable different to be detected.[25–27] Misclassification bias of exposure may arise as a result of these inaccuracies in spatial assignment of air pollution exposure. Some reasons this error occurs include: differences in average exposures of populations with individual exposures (such as due to differences in inhalation and uptake), spatial and temporal variability of $PM_{2.5}$ relative to data availability, and precision and accuracy of the instruments and devices used to quantify $PM_{2.5}$ concentrations.[28] This misclassification bias is more likely nondifferential and modifying the effects produced towards the null.[28,29]

4.6 Conclusion

In conclusion, the results of the present study provide no evidence to suggest smoking acts synergistically with $PM_{2.5}$ exposure on the incidence of T2DM. Further biologic mechanistic research and epidemiologic research using larger datasets is necessary to adequately elucidate on the true nature of this interaction. Finally, more complex models for interaction analyses may be needed to truly characterize the interaction.

References for Article 2

1. World Health Organization. Global report on diabetes. World Health Organization; 2016.
2. Thiering E, Heinrich J. Epidemiology of air pollution and diabetes. Trends Endocrinol Metab. 2015 Jul;26(7):384–94.
3. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol. 2011 Nov 8;8(4):228–36.
4. Turner MC, Cohen A, Jerrett M, Gapstur SM, Diver WR, Pope CA, et al. Interactions between cigarette smoking and fine particulate matter in the Risk of Lung Cancer Mortality in Cancer Prevention Study II. Am J Epidemiol. 2014 Dec 15;180(12):1145–9.
5. Samet J. Chapter 13: Combined Effects of Air Pollution with Other Agents. In: Straif K, Cohen A, editors. Air Pollution and Cancer. Lyon, France: World Health Organization; 2013.
6. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012 Apr 1;41(2):514–20.
7. Ahlbom A, Alfredsson L. Interaction: A word with two meanings creates confusion. Eur J Epidemiol. 2005;20(7):563–4.
8. Andersson T, Alfredsson L, Källberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol. 2005;20(7):575–9.
9. Rothman K, Lash T, Greenland S. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012.
10. Liang F, Yang X, Liu F, Li J, Xiao Q, Chen J, et al. Long-term exposure to ambient fine particulate matter and incidence of diabetes in China: A cohort study. Environ Int. 2019;126:568–75.
11. Chen Z, Peto R, Zhou M, Iona A, Smith M, Yang L, et al. Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. Lancet. 2015 Oct 10;386(10002):1447–56.
12. Chen X, Wu Z, Chen Y, Wang X, Zhu J, Wang N, et al. Risk score model of type 2 diabetes prediction for rural Chinese adults: the Rural Deqing Cohort Study. J Endocrinol Invest. 2017 Oct;40(10):1115–23.
13. Fu C, Chen Y, Wang F, Wang X, Song J, Jiang Q. High prevalence of hyperglycaemia and the impact of high household income in transforming rural China. BMC Public Health. 2011 Nov 12;11:862.
14. van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC, et al. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environ Sci Technol. 2016 Apr 5;50(7):3762–72.
15. Boys BL, Martin RV, van Donkelaar A, MacDonell RJ, Hsu NC, Cooper MJ, et al. Fifteen-Year Global Time Series of Satellite-Derived Fine Particulate Matter. Environ Sci Technol. 2014 Oct 7;48(19):11109–18.
16. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. Environ Health Perspect. 2015 Feb;123(2):135–43.

17. Li R, Chambless L. Test for additive interaction in proportional hazards models. Ann Epidemiol. 2007 Mar;17(3):227–36.

18. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. Epidem Methods. 2014 Dec 1;3(1):33–72.

19. de Mutsert R, Jager KJ, Zoccali C, Dekker FW. The effect of joint exposures: examining the presence of interaction. Kidney Int. 2009 Apr 1;75(7):677–81.

20. Pope CA, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, et al. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. Circulation. 2004 Jan 6;109(1):71–7.

21. Xu X, Wang L. Synergistic effects of air pollution and personal smoking on adult pulmonary function. Arch Environ Health. 1998 Feb;53(1):44–53.

22. Lin H, Guo Y, Kowal P, Airhihenbuwa CO, Di Q, Zheng Y, et al. Exposure to air pollution and tobacco smoking and their combined effects on depression in six low- and middle-income countries. Br J Psychiatry. 2017 Sep;211(3):157–62.

23. Cai L, Wang X-M, Fan L-M, Cui W-L, Golden AR. Socioeconomic disparities in prevalence and behaviors of smoking in rural Southwest China. BMC Public Health [Internet]. 2019 Aug 14 [cited 2020 Apr 27];19. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6694669/

24. Liu S, Zhang M, Yang L, Li Y, Wang L, Huang Z, et al. Prevalence and patterns of tobacco smoking among Chinese adult men and women: findings of the 2010 national smoking survey. J Epidemiol Community Health. 2017 Feb;71(2):154–61.

25. WHO Global Air Quality Guidelines Working Group on Risk of Bias Assessment. Risk of bias assessment instrument for systematic reviews informing WHO global air quality guidelines (2020). Copenhagen, Denmark: World Health Organization Regional Office for Europe; 2020 Feb.

26. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation. 2010 Jun 1;121(21):2331–78.

27. Brauer M. How much, how long, what, and where: air pollution exposure assessment for epidemiologic studies of respiratory disease. Proc Am Thorac Soc. 2010 May;7(2):111–5.

28. Lipfert FW, Wyzga RE. Statistical considerations in determining the health significance of constituents of airborne particulate matter. J Air Waste Manag Assoc. 1999;49(9):182–91.

29. Avery CL, Mills KT, Williams R, McGraw KA, Poole C, Smith RL, et al. Estimating Error in Using Residential Outdoor PM2.5 Concentrations as Proxies for Personal Exposures: A Meta-analysis. Environ Health Perspect. 2010 May;118(5):673–8.

Supplementary Material for Article 2

Supplementary Material 1. Sensitivity Analysis for First Quartile Cut-offs for Dichotomizing $PM_{2.5}$

Table S1.1. Baseline Exposure to Ambient Fine Particulate Matter Using First Quartile Cut-Off for Exposure, Overall and According to Gender

Baseline Characteristics	Overall				Male				Female			
	Total cohort		Cases of T2DM		All males		Cases of T2DM		All females		Cases of T2DM	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Total cohort	28,279	100.0	386	100.0	12,513	100.0	144	100.0	15,766	100.0	242	100.0
Smoking status												
Non-smokers	21,857	77.3	308	79.8	6,404	51.2	71	49.3	15,453	98.0	237	97.9
Smokers	6,422	22.7	78	20.2	6,109	48.8	73	50.7	313	2.00	5	2.1
$PM_{2.5}$ exposure												
Low (<47.17 $\mu g/m^3$)	7,189	25.4	94	24.3	3,281	26.2	36	25.0	3,908	24.8	58	24.0
High (≥47.17 $\mu g/m^3$)	21,090	74.6	292	75.7	9,232	73.8	108	75.0	11,858	75.2	184	76.0

The median concentration of baseline exposure to $PM_{2.5}$ (50.83 $\mu g/m^3$) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the main analysis. The 1st quartile concentration of baseline exposure to $PM_{2.5}$ (47.17 $\mu g/m^3$) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the sensitivity analysis.

All values of $PM_{2.5}$ are based on baseline exposure models unless otherwise stated.

Abbreviations: T2DM, type 2 diabetes mellitus; $PM_{2.5}$, ambient fine particulate matter.

Table S1.2. Incidence Rates of Type 2 Diabetes Mellitus by Baseline Exposure to Ambient Fine Particulate Matter Using First Quartile Cut-Off for Exposure, Overall and According to Gender

Baseline Characteristics	Overall				Males				Females			
	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI
Total cohort	386	118,117.3	3.27	2.96, 3.61	144	52,785.9	2.73	2.32, 3.21	242	65,331.5	3.70	3.27, 4.20
Smoking status												
Non-smokers	308	88,459.4	3.48	3.11, 3.89	71	24,617.2	2.88	2.29, 3.64	237	63,842.2	3.71	3.27, 4.22
Smokers	78	29,657.9	2.63	2.11, 3.28	73	28,168.6	2.59	2.06, 3.26	5	1,489.3	3.36	1.40, 8.08
$PM_{2.5}$ exposure												
Low (<47.17 μg/m^3)	94	36,679.4	2.56	2.09, 3.14	36	16,958.8	2.12	1.53, 2.94	58	19,720.7	2.94	2.27, 3.80
High (≥47.17 μg/m^3)	292	81,437.9	3.59	3.20, 4.02	108	35,827.1	3.01	2.50, 3.64	184	45,610.8	4.03	3.49, 4.66

The 1st quartile concentration of baseline exposure to $PM_{2.5}$ (47.17 μg/m^3) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the sensitivity analysis.

All values of $PM_{2.5}$ are based on baseline exposure models unless otherwise stated.

Abbreviations: T2DM, type 2 diabetes mellitus; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter.

Table S1.3. Multiplicative Effect Modification of Smoking Status and Baseline Exposure to Ambient Fine Particulate Matter on Incidence of Type 2 Diabetes Mellitus Using First Cut-Off for Exposure, Overall

	Cases of T2DM	Unadjusted model			Parsimoniously-adjusted model			Fully-adjusted model		
		HR	95% CI	p-value for interaction	HR	95% CI	p-value for interaction	HR	95% CI	p-value for interaction
Overall	**386**			0.2551			0.3384			0.3458
Non-smokers										
Low $PM_{2.5}$ (<47.17 μg/m^3)	77	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥47.17 μg/m^3)	231	1.24	0.95, 1.60		1.23	0.94, 1.59		1.28	0.98, 1.68	
Smokers										
Low $PM_{2.5}$ (<47.17 μg/m^3)	17	1.00	Ref		1.00	Ref		1.00	Ref	
High $PM_{2.5}$ (≥47.17 μg/m^3)	61	1.75	1.02, 2.99		1.64	0.96, 2.81		1.71	1.00, 2.94	

The 1st quartile concentration of baseline exposure to $PM_{2.5}$ (47.17 μg/m^3) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the sensitivity analysis.
Abbreviations: HR, hazard ratio, per 10-μg/m^3 increase in $PM_{2.5}$; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter.

Table S1.4. Joint Effect of Smoking Status and Baseline Exposure to Ambient Fine Particulate Matter on Incidence of Type 2 Diabetes Mellitus Using First Quartile Cut-Off for Exposure, Overall

	Cases of T2DM	Unadjusted model		Parsimoniously-adjusted model		Fully-adjusted model	
		HR	95% CI	HR	95% CI	HR	95% CI
Overall	**386**						
Non-smokers and low $PM_{2.5}$	138	1.00	Ref	1.00	Ref	1.00	Ref
Independent effect of smoking	17	0.60	0.35, 1.01	0.66	0.39, 1.12	0.82	0.47, 1.43
Independent effect of high $PM_{2.5}$	231	1.24	0.95, 1.60	1.22	0.94, 1.59	1.32	1.01, 1.72
Joint effect of smoking and high $PM_{2.5}$	61	1.04	0.74, 1.46	1.08	0.77, 1.52	1.46	0.99, 2.16

The 1st quartile concentration of baseline exposure to $PM_{2.5}$ (47.17 µg/m³) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the sensitivity analysis.

Abbreviations: T2DM, type 2 diabetes mellitus; HR, hazard ratio; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter.

Table S1.5. Measure of Additivity for Smoking Status and Baseline Exposure to Ambient Fine Particulate Matter on Incidence of Type 2 Diabetes Mellitus Using First Cut-Off for Exposure, Overall

Measure of Additivity	Overall	
	Point estimate	95% CI
Unadjusted model		
RERI	-0.064	-0.31, 0.18
AP	-0.083	-0.67, 0.51
SI	1.38	0.50, 3.79
Parsimoniously-adjusted model		
RERI	-0.0093	-0.31, 0.29
AP	-0.011	-0.55, 0.53
SI	1.08	0.12, 9.94
Fully-adjusted model		
RERI	0.44	-0.47, 1.35
AP	0.28	-0.13, 0.69
SI	4.25	0.32, 56.78

The 1st quartile concentration of baseline exposure to $PM_{2.5}$ (47.17 µg/m³) was used as cut-off for low vs. high $PM_{2.5}$ exposure for the sensitivity analysis.

Abbreviations: CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; RERI, relative excess risk for interaction; AP, attributable proportion of interaction; SI, synergy index.

_____ End of Article 2 _____

CHAPTER 5. Evaluating Roles of $PM_{2.5}$ Constituents in Association with Incidence of T2DM (Article 3)

Abstract for Article 3

Background: Although ambient fine particulate matter ($PM_{2.5}$) shows evidence of an association with the incidence of type 2 diabetes mellitus (T2DM), the specific constituents of $PM_{2.5}$ that lead to detrimental health effects are not known. $PM_{2.5}$ is commonly represented as total mass in epidemiologic studies of long-term associations with the development of T2DM; however, it may not be the best choice to represent the true adverse health effects of $PM_{2.5}$ on the incidence of T2DM.

Objective: The purpose of this article is to identify which $PM_{2.5}$ constituents may play major roles in the effect of long-term exposure to $PM_{2.5}$ on the incidence of T2DM in rural adults in China.

Methods: Ground-level $PM_{2.5}$ concentrations were approximated from satellite-derived ambient $PM_{2.5}$ estimates. In addition to $PM_{2.5}$, six of its main constituents, namely black carbon (BC), nitrate (NO_3^-), organic carbon (OC), sulfate (SO_4^{2-}), soil dust, and ammonium (NH_4^+), were identified by partitioning $PM_{2.5}$ mass into compositions using the Goddard Earth Observing System Chemical Model (GEOS-CHEM) simulation, a chemical transport model. The Rural Deqing Cohort data were used to obtain details on T2DM over the period from 2006 to 2014 and followed up until December 31, 2018. Three models of $PM_{2.5}$ constituents with T2DM were implemented: 1) the basic single-constituent, 2) the constituent-$PM_{2.5}$ joint model, and 3) the constituent-residual model. Each method of modelling used the Cox proportional hazards model to examine the effect of $PM_{2.5}$ constituents on incidence of T2DM. A p-value of 0.05 was chosen as the cut-off for considering results statistically significant. Associations were presented as hazard ratios (HR) with 95% confidence intervals (CI) per interquartile (IQR) increase in constituent.

Results: Based on the results of the parsimoniously-adjusted constituent residual model, OC, SO_4^{2-}, and NH_4^+ presented the strongest associations with T2DM incidence but with varying statistical significance depending on the exposure model.

Conclusions: There was insufficient evidence to indicate a particular constituent was more strongly associated with T2DM. Different models provided only modestly consistent results and therefore no conclusive interpretation is possible. Differing opinions on suitability of model choices and personal concerns of the impact of differing exposure assignment limits confidence in conclusions. Development of more complex constituent modelling approaches is needed to aid future studies interested in characterizing the relationship between $PM_{2.5}$ constituents and T2DM.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AOD	Aerosol optical depth
BC	Black carbon
BMI	Body mass index
CI	Confidence interval
CTM	Chemical transport model
EPA	Environmental Protection Agency
GBD	Global Burden of Disease
GEOS-CHEM	Goddard Earth Observing System Chemical Model
HEI	Health Effects Institute
HNO_3	Nitric acid
HR	Hazard ratio
IQR	Interquartile range
ISA	Integrated Science Assessment
MAIAC	Multi-Angle Implementation of Atmospheric Correction
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer Dark Target
NH_3	Ammonia
NH_4^+	Ammonium
NO_3^-	Nitrate
OC	Organic carbon
$PM_{2.5}$	Fine particulate matter
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor Deep Blue
SO_2	Sulfur dioxide
SO_4^{2-}	Sulfate
T2DM	Type 2 diabetes mellitus

5.1 Introduction

In a previous article (**Article 1**), results indicated that long-term exposure to fine particulate matter ($PM_{2.5}$) was associated with an increased incidence of type 2 diabetes mellitus (T2DM). The results were consistent with the conclusions of multiple other cohort studies from different regions of the world, thereby indicating that despite differing baseline exposure concentrations, sources, and human underlying risks, T2DM is consistently associated with long-term exposure to $PM_{2.5}$. Most of epidemiologic research of the relationship between $PM_{2.5}$ and adverse health effects is based on the measures of its mass. However, $PM_{2.5}$ is made up of a complex mixture of particles from differing sources with different chemical components.[1] Major constituents of $PM_{2.5}$ include black carbon (BC), nitrate (NO_3^-), organic carbon (OC), sulfate (SO_4^{2-}), soil dust, and ammonium (NH_4^+).[2]

Researchers have raised the concern that $PM_{2.5}$ mass itself is unlikely responsible for air pollution's adverse health effects alone,[3] therefore making the study of the effects of chemical composition of $PM_{2.5}$ on health essential.[4,5] In an attempt to identify which components of $PM_{2.5}$ are more toxic as well as the sources of pollution responsible for adverse health effects, many short-term and long-term studies have been published,[3,5] and some improperly attempt to model associations between $PM_{2.5}$ constituents and health outcomes using the same statistical methods that are used in studies of $PM_{2.5}$ mass.[6]

It is important to know which constituents are potentially more toxic, as policies to inform air quality strategies result in changes in the chemical composition of ambient air pollution.[7] For example, China's Air Pollution Prevention and Control Action Plan, implemented in 2013, has resulted not only in decreases in $PM_{2.5}$ in certain regions of China, but also varying reductions in constituents. Research has shown that SO_4^{2-} has had the largest reduction, and NO_3^- has had the lowest reduction as a result of the Action Plan.[8] Reduction in SO_4^{2-} are in line with the reduced sulfur dioxide (SO_2) emissions, driven largely

by the industrial sector. However, reduction in NO_3^-, largely driven by the power sector, is limited because of the chemical relationship with other constituents.[8] The decrease in SO_2 emissions but stable ammonia (NH_3) emissions allowed for greater formation of NO_3^- from nitric acid (HNO_3).[8] Truly, understanding which constituents of $PM_{2.5}$ are responsible for the most harmful effects is an essential and critical policy issue.[9]

In the current analysis, specific $PM_{2.5}$ constituents were evaluated to determine which might be responsible for the association between $PM_{2.5}$ mass and the incidence of T2DM in rural Chinese adults.

5.2 Methods

Study Design and Population

Study participants were identified from the Rural Deqing Cohort Study, with the initial purpose of creating a T2DM rural risk assessment tool.[10] Participants were adult residents aged 18 to 64 years, randomly cluster sampled from eight rural townships, including 65 villages in Deqing County, Zhejiang Province, China. Adults without local residency (i.e., temporary workers or university students not living in the country during the study period) were excluded from the analytical dataset. Participants included in the present analytical dataset were free of T2DM at baseline when they entered into the cohort during the period from 2006 to 2014 and were followed up until December 31, 2018 to identify new cases of T2DM. Details on the study design and population can be found in a previous article (**Article 1**).

Exposure Assessment

Satellite-derived ambient $PM_{2.5}$ estimates were used to approximate ground-level concentrations of $PM_{2.5}$ using the V4.CH.02 product of the Dalhousie University Atmospheric

Composition Analysis Group. In addition to $PM_{2.5}$, six of its main constituents, including organic carbon (OC), black carbon (BC), sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+) and soil dust, were evaluated using the V4.CH.02 product of the Dalhousie University Atmospheric Composition Analysis Group.[2] The dataset combines satellite retrievals and simulations of aerosol optical depth (AOD) from multiple sources [Multi-angle Imaging SpectroRadiometer (MISR), Moderate Resolution Imaging Spectroradiometer (MODIS) Dark Target, MODIS and Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) Deep Blue, and MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC)] with geophysically-based simulated relationships between AOD and near-surface $PM_{2.5}$ concentrations. The Goddard Earth Observing System Chemical Model (GEOS-CHEM), a chemical transport model (CTM), was used to partition $PM_{2.5}$ mass into its major constituents (BC, NO_3^-, OC, SO_4^{2-}, soil dust, and NH_4^+)[2,11,12] using meteorological datasets, emission inventories, and equations representing atmospheric physics and chemistry.[13] Monthly geographically weighted regressions at a 0.01° × 0.01° resolution from 2014 to 2016 were used to incorporate ground-based observations from China's recently expanded monitoring network.[14] The resulting estimates of annual average $PM_{2.5}$ concentrations were consistent with out of sample, cross-validation observations ($R^2 = 0.78$). $PM_{2.5}$ concentrations prior to this time period were based on the relative changes of Atmospheric Composition Analysis Group's long-term product.[2,12]

Linkage between estimates of annual average $PM_{2.5}$ concentrations and individual participant exposure was done by using each participant's home address at baseline on a 1 × 1 km spatial scale. $PM_{2.5}$ exposure was assigned as the baseline concentration as a surrogate for overall exposure in the analytical dataset. Participants with missing exposure data were excluded from the analytical dataset. $PM_{2.5}$ exposure assigned as the 3-year, time-fixed average and the yearly time-varying concentration as surrogates for overall exposure were additionally used in sensitivity analyses.

Outcome and Covariates

Participants who reported having physician-diagnosed T2DM at baseline were excluded from the analytical dataset. Incident cases of T2DM were identified from the testing results of blood samples that were collected as part of physical examinations over the follow-up period, self-reported diagnosis of T2DM, or use of anti-diabetic medication. A new case of T2DM identified from a blood sample was defined as an FPG of 7.0 mmol/L or higher following an overnight fast of at least 8 hours. T2DM cases were ascertained from Deqing electronic health records up until December 31, 2018.

Demographic and behavioural covariates, such as education level and smoking status, were determined from information collected in the baseline questionnaire. Anthropometric covariates, such as height and weight, used to calculate body mass index (BMI), were determined from the physical examination at baseline. Details on the included covariates and categorical coding choices of covariates are presented in a previous article (**Article 1**).

5.3 Statistical Analysis

Main Analysis

Descriptive statistics, such as annual average concentrations over the study period, were presented for $PM_{2.5}$ and its constituents (i.e. BC, NO_3^-, OC, SO_4^{2-}, soil dust, and NH_4^+). Spatial and temporal distributions over the study region during the study period were additionally examined for $PM_{2.5}$ and its constituents. Histograms representing the distributions of exposure to $PM_{2.5}$ and its constituents for cohort participants were generated to demonstrate whether a normal distribution was observed for each measure. The number of events and incidence rates of T2DM for quartiles of exposure to $PM_{2.5}$ and its constituents, were presented to assess whether incidence rates varied by

quartile categories of exposure. Correlations between $PM_{2.5}$ and its constituents were assessed using the Pearson correlation coefficient.

Cox proportional hazards model was used to examine the association between long-term exposure to $PM_{2.5}$ and incidence of T2DM. $PM_{2.5}$ and its constituents were assigned to each participant as the baseline concentration, and as 3-year averages and yearly time-varying exposures in sensitivity analyses. To evaluate the association between $PM_{2.5}$ constituents and incidence of T2DM, three modelling strategies were implemented (**Tables S1**): the basic single-constituent model, the constituent-$PM_{2.5}$ joint model, and the constituent-residual model, as previously implemented in other research.[6,15,16] All methods include each constituent individually into the model.

In the basic single-constituent model, each constituent is individually incorporated into the basic model. The model is represented as:

$$g(\mu) = \beta 0 + \beta 1(\text{constituent}) + [\gamma' X]$$

where the dependent variable for the health outcome (μ) is a function of the independent variables for constituent, total $PM_{2.5}$ mass, and a matrix of other covariates [γ'X].[16]

In the constituent-$PM_{2.5}$ joint model, each constituent is individually incorporated into the model including adjustment for $PM_{2.5}$ mass to account for the potential confounding effect $PM_{2.5}$ may have on constituents. The model is represented as:

$$g(\mu) = \beta 0 + \beta 1(\text{constituent}) + \beta 2(PM_{2.5}) + [\gamma' X]$$

In the constituent-residual model, the residual of each constituent from a linear regression model between $PM_{2.5}$ mass and that constituent, is individually incorporated into the basic model, replacing the constituent. In theory, this constituent residual represents the independent contribution of that

constituent to $PM_{2.5}$'s effects on T2DM risk, whilst eliminating the issue of overadjustment and collinearity with other constituents.[6,15] The model is represented as:

$$g(\mu) = \beta 0 + \beta 1(\text{residual}) + [\gamma' X]$$

where residual is identified from the linear regression between constituent and total $PM_{2.5}$ mass.

For each modelling method, the association is presented as hazard ratios (HR) with 95% confidence intervals (CI) per interquartile range (IQR) increase in constituent exposure. All models are presented as unadjusted and parsimoniously-adjusted models. The parsimoniously-adjusted model includes only covariates chosen *a priori* that additionally showed statistical evidence of confounding for the association between $PM_{2.5}$ exposure and incidence of T2DM, minimizing the potential issues of overadjustment.[17–19] Further details on the identification of potential confounding covariates for parsimonious adjustment can be found in a previous article (**Article 1**).

All statistical tests were two-sided and p-value of less than 0.05 was considered statistically significant. All analysis was performed using SAS 9.4.

Sensitivity Analysis and Model Assumptions

In the main analysis, exposure to $PM_{2.5}$ and its constituents were assigned as the concentrations in the baseline year of entry. To assess whether results were consistent regardless of exposure metric, sensitivity analyses were conducted by repeating the main analysis using exposure assigned as 3-year, time-fixed averages and yearly time-varying concentrations.

The proportional hazards assumption was assessed by implementing the <ASSESS> statement in SAS, for which a p-value less than 0.05 indicated evidence of disproportionate hazards.[20,21]

5.4 Results

Descriptive Statistics

The descriptive statistics on the average concentrations of baseline $PM_{2.5}$ and its constituents are presented in **Table 1**. The average baseline concentration with standard deviations (SD) of $PM_{2.5}$, BC, NO_3^-, OC, SO_4^{2-}, soil dust, and NH_4^+, were 51.15 μg/m^3 (SD: 5.27 μg/m^3), 3.24 μg/m^3 (SD: 0.35 μg/m^3), 14.23 μg/m^3 (SD: 1.66 μg/m^3), 8.94 μg/m^3 (SD: 0.96 μg/m^3), 13.13 μg/m^3 (SD: 1.61 μg/m^3), 2.65 μg/m^3 (SD: 0.62 μg/m^3), and 8.96 μg/m^3 (SD: 0.86 μg/m^3), respectively. The descriptive statistics for 3-year average-assigned and time-varying $PM_{2.5}$ and its constituents are presented in **Table S4.1** and **Table S5.1**, respectively.

Table 1. Descriptive Statistics for Baseline Concentration of Ambient Fine Particulate Matter and Its Constituents in Deqing County

Baseline concentration (μg/m^3)	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
$PM_{2.5}$	51.14	5.27	39.42	47.17	50.83	56.08	82.92	8.92
BC	3.24	0.35	2.03	2.89	3.32	3.55	4.35	0.66
NO_3^-	14.23	1.66	6.31	13.07	14.80	15.52	19.10	2.45
OC	8.94	0.96	7.01	8.03	9.21	9.69	17.12	1.66
SO_4^{2-}	13.13	1.61	9.36	12.14	12.95	14.58	20.26	2.44
Soil dust	2.65	0.62	1.55	2.04	2.81	3.08	16.72	1.03
NH_4^+	8.96	0.86	6.39	8.38	8.98	9.68	11.93	1.29

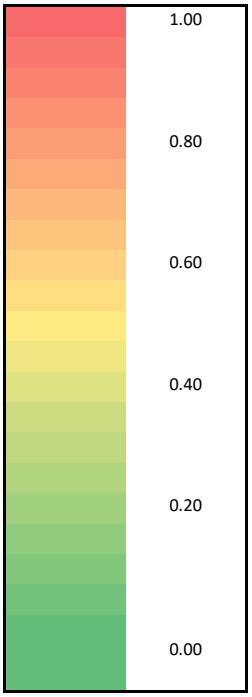
Abbreviations: SD, standard deviation; IQR, interquartile range; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Heat maps of average concentrations of $PM_{2.5}$ and its constituents in Deqing County, 2006 visually represent the spatial distribution over the study region on a 1 × 1 km grid (**Figures S2.1–S2.7**). Individual tables of descriptive statistics for $PM_{2.5}$ and its constituents for each year in the study period and graphs of average annual concentration over the study period demonstrate the temporal distribution (**Tables S3.1–S3.7**). Histograms of average annual concentrations of $PM_{2.5}$ and its constituents visually confirmed that the population distribution is normally distributed.

Correlations between baseline exposure to $PM_{2.5}$ and its constituents are presented **Figure 1**. There were high correlations amongst $PM_{2.5}$ and its constituents, ranging from 0.34 (between NO_3^- and SO_4^{2-}) to 0.98 (between $PM_{2.5}$ and NH_4^+). Expectantly, the strongest correlations were between $PM_{2.5}$ mass and each constituent individually, and less so for constituents with one another. Correlations for 3-year average exposure and time-varying exposure are presented in **Figure S4.1** and **Figure S5.1**, respectively.

Figure 1. Correlation between Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

	$PM_{2.5}$	BC	NO_3^-	OC	SO_4^{2-}	Soil dust	NH_4^+
$PM_{2.5}$	1	0.94	0.83	0.92	0.78	0.79	0.98
BC	0.94	1	0.96	0.92	0.56	0.63	0.90
NO_3^-	0.83	0.96	1	0.86	0.34	0.42	0.78
OC	0.92	0.92	0.86	1	0.53	0.77	0.83
SO_4^{2-}	0.78	0.56	0.34	0.53	1	0.75	0.85
Soil dust	0.79	0.63	0.42	0.77	0.75	1	0.73
NH_4^+	0.98	0.90	0.78	0.83	0.85	0.73	1



Abbreviations: $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

The number of new cases of T2DM out of the total cohort is presented according to quartiles of baseline exposure to $PM_{2.5}$ and its constituents (**Table 2**). Additionally, incidence rates of T2DM are presented according to quartiles of baseline exposure to $PM_{2.5}$ and its constituents (**Table 3**). Incidence rates according to quartiles of exposure suggested potential positive associations between T2DM and BC, NO_3^-, and OC. Further investigation is presented in the analytical models.

Table 2. Distribution of Quartiles of Baseline Concentration of Ambient Fine Particulate Matter and Its Constituents, According to Diabetes Status at End of Study

Baseline concentration (μg/m^3)	Total cohort		New cases of T2DM		Participants without T2DM at end of follow-up	
	No.	%	No.	%	No.	%
Overall	28,279	100.0	386	100.0	27,893	100.0
$PM_{2.5}$						
1st quartile (<47.17)	7,189	25.4	94	24.4	7,095	25.4
2nd quartile (47.17–50.82)	6,895	24.4	119	30.8	6,776	24.3
3rd quartile (50.83–56.07)	6,837	24.2	84	21.8	6,753	24.2
4th quartile (≥56.08)	7,358	26.0	89	23.1	7,269	26.1
BC						
1st quartile (<2.89)	5,620	19.9	109	28.2	5,511	19.8
2nd quartile (2.89–3.31)	8,634	30.5	104	26.9	8,530	30.6
3rd quartile (3.32–3.54)	7,010	24.8	110	28.5	6,900	24.7
4th quartile (≥3.55)	7,015	24.8	63	16.3	6,952	24.9
NO_3^-						
1st quartile (<13.07)	7,159	25.3	142	36.8	7,017	25.2
2nd quartile (13.07–14.79)	6,843	24.2	67	17.4	6,776	24.3
3rd quartile (14.80–15.51)	7,414	26.2	96	24.9	7,318	26.2
4th quartile (≥15.52)	6,863	24.3	81	21.0	6,782	24.3
OC						
1st quartile (<8.03)	6,946	24.6	134	34.7	6,812	24.4
2nd quartile (8.03–9.20)	7,750	27.4	87	22.5	7,663	27.5
3rd quartile (9.21–9.68)	5,983	21.2	76	19.7	5,907	21.2
4th quartile (≥9.69)	7,600	26.9	89	23.1	7,511	26.9
SO_4^{2-}						
1st quartile (<12.14)	6,980	24.7	80	20.7	6,900	24.7
2nd quartile (12.14–12.94)	7,047	24.9	92	23.8	6,955	24.9
3rd quartile (12.95–14.57)	7,110	25.1	133	34.5	6,977	25.0
4th quartile (≥14.58)	7,142	25.3	81	21.0	7,061	25.3
Soil dust						
1st quartile (<2.04)	6,742	23.8	75	19.4	6,667	23.9
2nd quartile (2.04–2.80)	7,774	27.5	141	36.5	7,633	27.4
3rd quartile (2.81–3.07)	6,870	24.3	79	20.5	6,791	24.4
4th quartile (≥3.08)	6,893	24.4	91	23.6	6,802	24.4
NH_4^+						
1st quartile (<8.38)	6,931	24.5	98	25.4	6,833	24.5
2nd quartile (8.38–8.97)	7,034	24.9	112	29.0	6,922	24.8
3rd quartile (8.98–9.68)	7,250	25.6	109	28.2	7,141	25.6
4th quartile (≥9.69)	7,064	25.0	67	17.4	6,997	25.1

Abbreviations: $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Table 3. Incidence Rates of Type 2 Diabetes Mellitus, According to Quartiles of Baseline Concentrations of Ambient Fine Particulate Matter and Its Constituents

Baseline concentration (μg/m^3)	Cases of T2DM	Person-years of follow-up	Incidence per 1,000 person-years	95% CI
Overall	386	118,117.3	3.27	2.96, 3.61
$PM_{2.5}$				
1st quartile (<47.17)	94	36,679.4	2.56	2.09, 3.14
2nd quartile (47.17–50.82)	119	38,146.4	3.12	2.61, 3.73
3rd quartile (50.83–56.07)	84	23,225.4	3.62	2.92, 4.48
4th quartile (≥56.08)	89	20,066.1	4.44	3.60, 5.46
BC				
1st quartile (<2.89)	109	39,764.8	2.74	2.27, 3.31
2nd quartile (2.89–3.31)	104	37,788.9	2.75	2.27, 3.34
3rd quartile (3.32–3.54)	110	25,784.4	4.27	3.54, 5.14
4th quartile (≥3.55)	63	14,779.2	4.26	3.33, 5.46
NO_3^-				
1st quartile (<13.07)	142	53,169.5	2.67	2.27, 3.15
2nd quartile (13.07–14.79)	67	25,895.1	2.59	2.04, 3.29
3rd quartile (14.80–15.51)	96	23,550.1	4.08	3.34, 4.98
4th quartile (≥15.52)	81	15,502.6	5.22	4.20, 6.50
OC				
1st quartile (<8.03)	134	46,536.1	2.88	2.43, 3.41
2nd quartile (8.03–9.20)	87	32,619.9	2.67	2.16, 3.29
3rd quartile (9.21–9.68)	76	19,798.4	3.84	3.07, 4.81
4th quartile (≥9.69)	89	19,162.9	4.64	3.77, 5.72
SO_4^{2-}				
1st quartile (<12.14)	80	25,284.8	3.16	2.54, 3.94
2nd quartile (12.14–12.94)	92	35,962.8	2.56	2.09, 3.14
3rd quartile (12.95–14.57)	133	31,146.9	4.27	3.60, 5.06
4th quartile (≥14.58)	81	25,722.9	3.15	2.53, 3.92
Soil dust				
1st quartile (<2.04)	75	22,314.6	3.36	2.68, 4.21
2nd quartile (2.04–2.80)	141	54,381.7	2.59	2.20, 3.06
3rd quartile (2.81–3.07)	79	22,816.9	3.46	2.78, 4.32
4th quartile (≥3.08)	91	18,604.2	4.89	3.98, 6.01
NH_4^+				
1st quartile (<8.38)	98	35,945.9	2.73	2.24, 3.32
2nd quartile (8.38–8.97)	112	39,328.0	2.85	2.37, 3.43
3rd quartile (8.98–9.68)	109	24,156.2	4.51	3.74, 5.44
4th quartile (≥9.69)	67	18,687.2	3.59	2.82, 4.56

Abbreviations: T2DM, type 2 diabetes mellitus; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Analytical Models

As described in the Methods, associations between constituents and the incidence of T2DM were evaluated by implementing three modelling strategies: the basic single-constituent model, the constituent-$PM_{2.5}$ joint model, and the constituent-residual model, as implemented in other research.[6,15,16] All HRs with 95% CIs were presented per IQR-increase in exposure and modelled unadjusted and parsimoniously-adjusted for potential confounding variables.

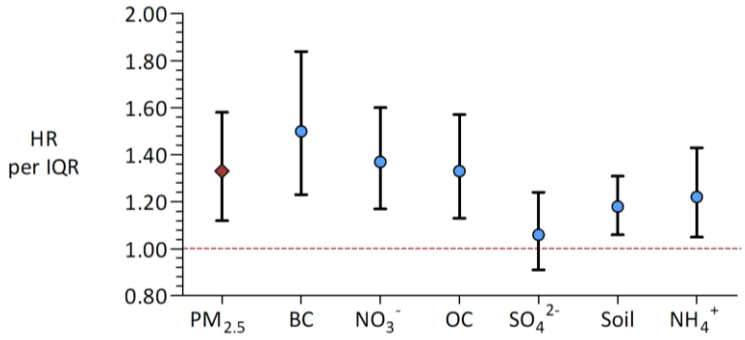
The results of the unadjusted single-constituent model are presented in **Table 4** and **Figure 2** and in **Table 5** and **Figure 3** for the parsimoniously-adjusted single-constituent model. In the unadjusted model, all the constituents except for SO_4^{2-} were significantly associated with incident T2DM. The HR for T2DM per IQR (0.66-$\mu g/m^3$) increase in BC was 1.50 (95% CI: 1.23, 1.84). The HR for T2DM per IQR (2.45-$\mu g/m^3$) increase in NO_3^- was 1.37 (95% CI: 1.17, 1.60). The HR for T2DM per IQR (1.66-$\mu g/m^3$) increase in OC was 1.33 (95% CI: 1.13, 1.57). The HR for T2DM per IQR (2.44-$\mu g/m^3$) increase in SO_4^{2-} was 1.06 (95% CI: 0.91, 1.24). The HR for T2DM per IQR (1.03-$\mu g/m^3$) increase in soil dust was 1.18 (95% CI: 1.06, 1.31). The HR for T2DM per IQR (1.29-$\mu g/m^3$) increase in NH_4^+ was 1.22 (95% CI: 1.05, 1.43). The results remained comparable when the model was parsimoniously-adjusted for age, BMI, and education; however, HRs for all constituents except soil dust were no longer statistically significant when adjusted for these potential confounders. For 3-year average exposure (**Tables S4.2–S4.3**), the association was positive and statistically significant only for BC, NO_3^-, and soil dust in the unadjusted model, and in the model parsimoniously adjusted for age, passing smoking exposure, vegetable consumption, history of disease, and occupation, only OC was statistically insignificant. The results from assigning exposure as time-varying were comparable (**Tables S5.2–S5.3**). For time-varying exposure, the association was similarly positively associated with all constituents, with only OC being statistically insignificant in the model parsimoniously adjusted for age, education, and occupation.

Table 4. Unadjusted Single-Constituent Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR	HR per IQR-increase	95% CI	p-value
1	$PM_{2.5}$	8.92	1.33	1.12, 1.58	0.0014
2	BC	0.66	1.50	1.23, 1.84	<.0001
3	NO_3^-	2.45	1.37	1.17, 1.60	0.0001
4	OC	1.66	1.33	1.13, 1.57	0.0006
5	SO_4^{2-}	2.44	1.06	0.91, 1.24	0.4511
6	Soil dust	1.03	1.18	1.06, 1.31	0.0018
7	NH_4^+	1.29	1.22	1.05, 1.43	0.0111

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase in exposure**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Figure 2. Unadjusted Single-Constituent Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents



*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
IQR: $PM_{2.5}$ (8.92 μg/m³); BC (0.66 μg/m³); NO_3^- (2.45 μg/m³); OC (1.66 μg/m³); SO_4^{2-} (2.44 μg/m³); soil dust (1.03 μg/m³); NH_4^+ (1.29 μg/m³).

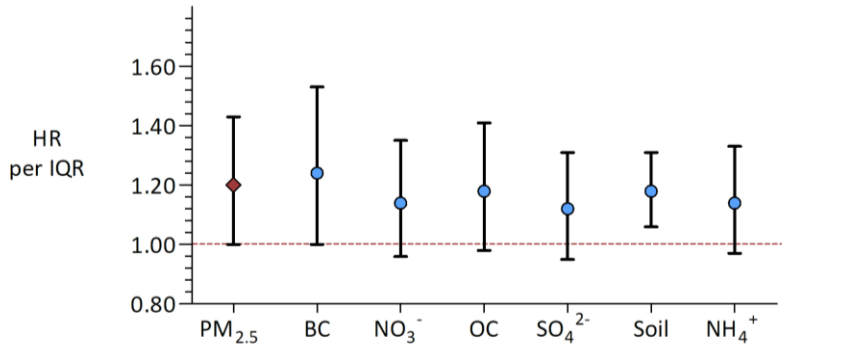
Table 5. Parsimoniously-Adjusted Single-Constituent Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR	HR per IQR-increase	95% CI	p-value
1	$PM_{2.5}$	8.92	1.20	1.00, 1.43	0.0523
2	BC	0.66	1.24	1.00, 1.53	0.0508
3	NO_3^-	2.45	1.14	0.96, 1.35	0.1314
4	OC	1.66	1.18	0.98, 1.41	0.0769
5	SO_4^{2-}	2.44	1.12	0.95, 1.31	0.1660

6	Soil dust	1.03	1.18	1.06, 1.31	0.0024
7	NH_4^+	1.29	1.14	0.97, 1.33	0.1111

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase in exposure**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, BMI, and education.

Figure 3. Parsimoniously-Adjusted Single-Constituent Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents



*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
IQR: $PM_{2.5}$ (8.92 μg/m³); BC (0.66 μg/m³); NO_3^- (2.45 μg/m³); OC (1.66 μg/m³); SO_4^{2-} (2.44 μg/m³); soil dust (1.03 μg/m³); NH_4^+ (1.29 μg/m³).
Covariates in the parsimoniously-adjusted model include age, BMI, and education.

The results of the unadjusted constituent-$PM_{2.5}$ joint model are presented in **Table 6** and **Figure 4** and in **Table 7** and **Figure 5** for the parsimoniously-adjusted constituent-$PM_{2.5}$ joint model. From the constituent-$PM_{2.5}$ joint model, the results can be interpreted as the impact of higher levels of the constituent, holding other constituents of $PM_{2.5}$ constant.[16] The associations between T2DM and constituents, adjusted for $PM_{2.5}$ mass but unadjusted for potential confounding covariates, varied between positively to negatively associated, with varying degrees of statistical significance. When parsimoniously-adjusted, results were more consistent and uniformly approximating the null (HR = 1.00). Only BC and soil were slightly positively associated with T2DM but without statistical significance. NH_4^+ was negatively associated with T2DM without statistical significance. For the 3-year average exposure sensitivity analysis (**Tables S4.4–S4.5**), BC and soil dust were similarly positively associated with T2DM without statistical significance, along with NO_3^- in the parsimoniously-adjusted constituent-

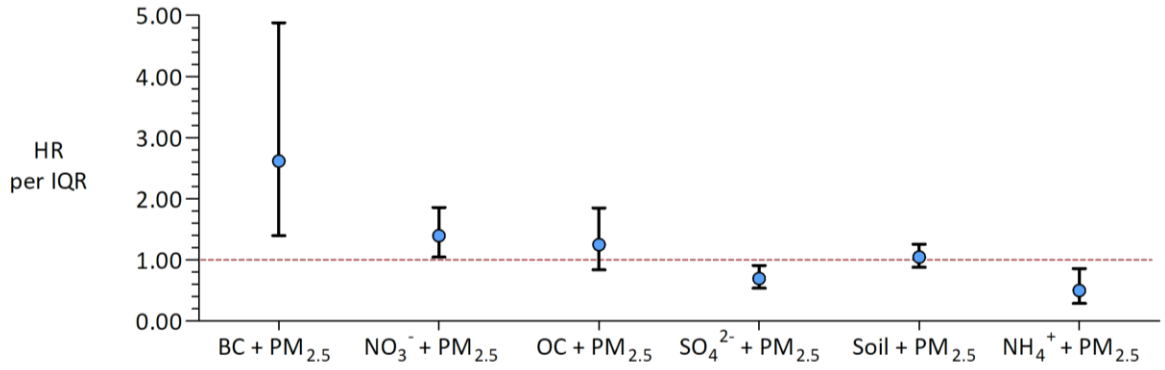
$PM_{2.5}$ joint model. OC, SO_4^{2-}, and NH_4^+ were negatively associated with T2DM without statistical significance. For time-varying exposure (**Tables S5.4–S5.5**), the constituent-$PM_{2.5}$ joint model's results were not comparable to those found for baseline and 3-year average exposure, as high correlations led to extremely wide CIs for BC, SO_4^{2-}, and NH_4^+.

Table 6. Unadjusted Constituent-$PM_{2.5}$ Joint Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC + $PM_{2.5}$	0.66	2.62	1.40, 4.88	0.0025
2	NO_3^-+ $PM_{2.5}$	2.45	1.40	1.05, 1.86	0.0223
3	OC + $PM_{2.5}$	1.66	1.25	0.84, 1.85	0.2784
4	SO_4^{2-} + $PM_{2.5}$	2.44	0.70	0.54, 0.91	0.0070
5	Soil dust + $PM_{2.5}$	1.03	1.05	0.88, 1.26	0.5772
6	NH_4^++ $PM_{2.5}$	1.29	0.50	0.29, 0.86	0.0121

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Figure 4. Unadjusted Constituent-$PM_{2.5}$ Joint Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents



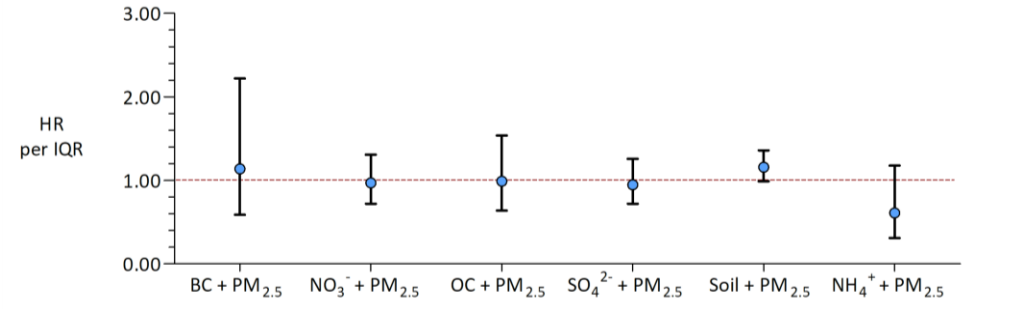
*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
IQR: BC (0.66 μg/m³); NO_3^- (2.45 μg/m³); OC (1.66 μg/m³); SO_4^{2-} (2.44 μg/m³); soil dust (1.03 μg/m³); NH_4^+ (1.29 μg/m³).

Table 7. Parsimoniously-Adjusted Constituent-$PM_{2.5}$ Joint Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC + $PM_{2.5}$	0.66	1.14	0.59, 2.22	0.6992
2	NO_3^- + $PM_{2.5}$	2.45	0.97	0.72, 1.31	0.8619
3	OC + $PM_{2.5}$	1.66	0.99	0.64, 1.54	0.9742
4	SO_4^{2-} + $PM_{2.5}$	2.44	0.95	0.72, 1.26	0.7272
5	Soil dust + $PM_{2.5}$	1.03	1.16	0.99, 1.36	0.0648
6	NH_4^+ + $PM_{2.5}$	1.29	0.61	0.31, 1.18	0.1425

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include $PM_{2.5}$, age, BMI, and education.

Figure 5. Parsimoniously-Adjusted Constituent-$PM_{2.5}$ Joint Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents



*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
IQR: BC (0.66 μg/m³); NO_3^- (2.45 μg/m³); OC (1.66 μg/m³); SO_4^{2-} (2.44 μg/m³); soil dust (1.03 μg/m³); NH_4^+ (1.29 μg/m³).
Covariates in the parsimoniously-adjusted model include age, BMI, and education.

The results of the unadjusted constituent residual model are presented in **Table 8** and **Figure 6** and in **Table 9** and **Figure 7** for the parsimoniously-adjusted constituent residual model. From the constituent residual model, the results can be interpreted as the risk associated with higher levels of the constituent while holding $PM_{2.5}$ mass constant, accounting for the correlation amongst constituents and $PM_{2.5}$.[16] In the unadjusted constituent residual model, SO_4^{2-}, soil dust, NH_4^+ were positively associated with T2DM, with SO_4^{2-} and NH_4^+ presenting statistical significance. All other constituents were not statistically significant and hovered closely around the null hypothesis (HR = 1.00). In the

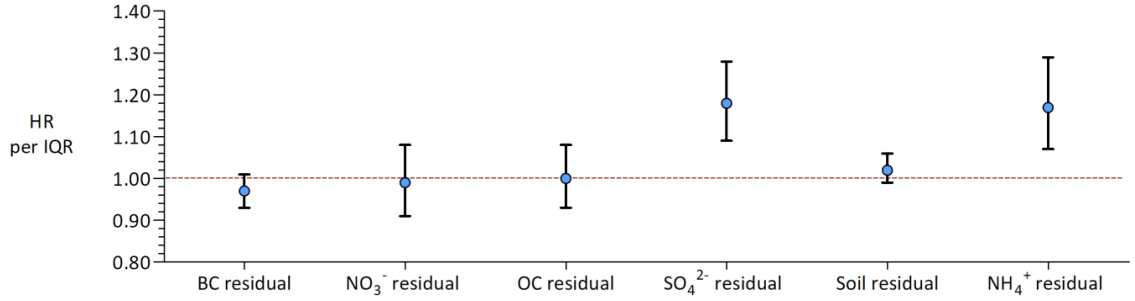
parsimoniously-adjusted constituent residual model, no constituent was statistically significant. Similar to the main analysis, the unadjusted constituent residual model using 3-year average exposure (**Tables S4.6–S4.7**) also resulted in positive and statistically significant $SO_4^{2-}$and NH_4^+, along with OC. In the parsimoniously-adjusted constituent residual model using 3-year average exposure, SO_4^{2-}, NH_4^+, and OC remained positively associated with statistical significance. Comparatively, in the unadjusted constituent residual model using time-varying exposure (**Tables S5.6–S5.7**), OC and soil dust were positively associated with statistical significance, and remaining constituents were positively associated without statistical significance. In the parsimoniously-adjusted constituent residual model using time-varying exposure, OC and soil dust remained statistically significant.

Table 8. Unadjusted Constituent Residual Model Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC residual	0.66	0.97	0.93, 1.01	0.1142
2	NO_3^- residual	2.45	0.99	0.91, 1.08	0.8686
3	OC residual	1.66	1.00	0.93, 1.08	0.9434
4	SO_4^{2-} residual	2.44	1.18	1.09, 1.28	<.0001
5	Soil dust residual	1.03	1.02	0.99, 1.06	0.2097
6	NH_4^+ residual	1.29	1.17	1.07, 1.29	0.0005

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Figure 6. Unadjusted Constituent Residual Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents



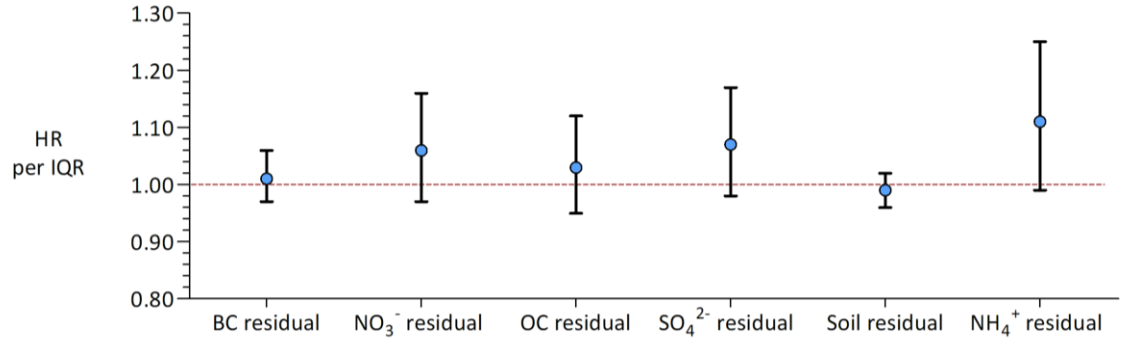
*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
IQR: BC (0.66 μg/m³); NO_3^- (2.45 μg/m³); OC (1.66 μg/m³); SO_4^{2-} (2.44 μg/m³); soil dust (1.03 μg/m³); NH_4^+ (1.29 μg/m³).

Table 9. Parsimoniously-Adjusted Constituent Residual Model Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC residual	0.66	1.01	0.97, 1.06	0.5608
2	NO_3^- residual	2.45	1.06	0.97, 1.16	0.1714
3	OC residual	1.66	1.03	0.95, 1.12	0.4344
4	SO_4^{2-} residual	2.44	1.07	0.98, 1.17	0.1462
5	Soil dust residual	1.03	0.99	0.96, 1.02	0.5219
6	NH_4^+ residual	1.29	1.11	0.99, 1.25	0.0624

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, BMI, and education.

Figure 7. Parsimoniously-Adjusted Constituent Residual Models Using Baseline Exposure to Ambient Fine Particulate Matter and Its Constituents



*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
IQR: BC (0.66 μg/m³); NO_3^- (2.45 μg/m³); OC (1.66 μg/m³); SO_4^{2-} (2.44 μg/m³); soil dust (1.03 μg/m³); NH_4^+ (1.29 μg/m³).
Covariates in the parsimoniously-adjusted model include age, BMI, and education.

In summary, the parsimoniously-adjusted constituent residual models suggested strongest associations for NO_3^-, SO_4^{2-}, and NH_4^+ when using baseline exposure, SO_4^{2-}, NH_4^+, and OC when using 3-year average exposure, NO_3^-, OC, and SO_4^{2-} when using time-varying exposure. Only SO_4^{2-} and NH_4^+ in 3-year average model and OC in the time-varying exposure model resulted in statistically significant

associations. In theory, the final presented model (i.e., the constituent residual model), best characterizes the association between a health outcome and $PM_{2.5}$ constituents.[6,15] Based on differing exposure assignment methods, conclusions regarding which constituents of $PM_{2.5}$ are most likely associated with increased incidence of T2DM would also differ.

Model Assumptions

The test for proportional hazards assumption for $PM_{2.5}$ and each constituent individually are presented in (**Figures S6.1–S6.7**)**.** Using the <ASSESS> statement in SAS, p-value less than 0.05 signifies disproportionate hazards. All constituents were not statistically significant, except for SO_4^{2-} with a p-value of 0.0110 (**Figure S6.5**), suggesting that for SO_4^{2-}, the proportional hazards assumption may not hold.

5.5 Discussion

Summary

Based on the assumption that the constituent residual model best represents the independent effect of $PM_{2.5}$ constituents on incidence of disease,[15] as well as the assumption that exposure assigned at baseline is a sufficient proxy for overall exposure,[22] OC, SO_4^{2-}, and NH_4^+ may be suggested as the constituents of $PM_{2.5}$ most likely to be associated with the incidence of T2DM in rural Chinese adults. However, similar to other studies on disease outcome and $PM_{2.5}$ constituents,[3] conclusions in the present study are not made with a great deal of confidence. Given the varying opinions on the best choice of analytical model (i.e. single-constituent, constituent-$PM_{2.5}$ joint, and constituent residual

model) and exposure assignment (i.e. baseline, 3-year average, and yearly time-varying exposure), confidence in these results is greatly limited.

Comparison to Other Research

Although there are several studies that have assessed the effects of long-term exposure to $PM_{2.5}$ constituents on all-cause, respiratory, and cardiovascular mortality,[23] no studies currently assess the association with T2DM. More studies have assessed the association between short-term exposure to $PM_{2.5}$ constituents and adverse health effects.[24–26] but there is insufficient literature on long-term exposure to $PM_{2.5}$ constituents.[26] In a systematic review with meta-analysis, the constituents of $PM_{2.5}$ with the most consistent evidence were BC and OC for all-cause, and BC, OC and NO_3^- for cardiovascular causes.[23] Authors additionally indicated that evidence suggested short-term exposure to NO_3^- and SO_4^{2-} were associated with cardiovascular mortality, and that NO_3^- was additionally linked to respiratory morbidity. However, authors concluded that there was insufficient evidence to address long-term exposure and adverse health effects.[23]

The US Environmental Protection Agency (EPA) published a 2019 Integrated Science Assessment (ISA) for particulate matter, including $PM_{2.5}$ and its constituents, and concluded that "The results of these studies confirm and further support the conclusion of the 2009 particulate matter ISA that many $PM_{2.5}$ components and sources are associated with many health effects and that the evidence does not indicate that any one source or component is consistently more strongly related with health effects than $PM_{2.5}$ mass".[27] In their literature search, the EPA did not identify any studies on the association between $PM_{2.5}$ constituents and T2DM, and very limited literature on metabolic effects overall.

Biologic Plausibility

Toxicological studies on BC have suggested that BC may not be a major toxic constituent of $PM_{2.5}$ but may serve as a universal carrier of toxic combustion-derived constituents (e.g. metals) to target regions of the human body. $SO_4^{2-}$and NO_3^- are secondary inorganic particles formed from gaseous primary pollutants and are highly soluble. Their high solubility and abundance in the human body has led researchers to suggest that they may be the least toxic of the $PM_{2.5}$ constituents. Furthermore, although epidemiological evidence suggests a potential association between $SO_4^{2-}$and adverse health effects, toxicological evidence provides little support for a causal relationship. However, it is possible that cations associated with SO_4^{2-} and NH_4^+ (e.g. transition metals), which are toxic, may be the source of the strong associations between SO_4^{2-} and adverse health effects seen in epidemiologic studies.[28]

Sources

Although the present study is limited in its lack of data on potential emission sources of $PM_{2.5}$, the constituents may help allude to possible sources. For instance, SO_4^{2-} and NO_3^- are secondary aerosols that may additionally serve as marker elements for automobile gasoline and diesel. SO_4^{2-} may additionally indicate oil and coal burning sources. BC and OC may suggest biomass burning or automobile diesel.[29] Insights from source apportionment studies on sources of $PM_{2.5}$ in China suggest that fossil fuel and industrial sources account for the majority of SO_4^{2-} fossil fuel and industrial sources followed closely behind by transportation account for the majority of NO_3^-, and agriculture for NH_4^+.[30] Dust, understandably, largely comes from the massive industrial production and multitude of infrastructure projects required to support China's economic growth.[30] BC and OC emissions are largely due to the burning of coal and biofuel in the residential sector for cooking and heating in rural homes,

and in the industrial sector due to poorly controlled coal-fired boilers and furnaces in less regulated regions, such as the rural industry.[31]

Burning of coal and other fuels from industrial and power plants, transportation, residential burning of biomass, open burning of agriculture fields, and residential burning of coal for cooking and heating are some of China's major sources of $PM_{2.5}$.[32,33] A special report from the Global Burden of Disease (GBD) Maps Working Group published from the Health Effects Institute (HEI) indicated that in Zhejiang province, where Deqing County resides, the major sources of ambient $PM_{2.5}$ included industrial sources from coal emissions, biomass combustion, transport, industrial sources of non-coal emissions, and power plants.[13] Moreover, although China has made massive changes and improvements in regulating $PM_{2.5}$ levels in urban areas, the current regulations completely ignore the necessity of equally stringent enforcement in rural areas as well.[33]

Limitations

One defining issue in the present study is the lack of co-pollutant data, such as nitric oxides or sulfur. This leads to challenges in interpreting the results as being from $PM_{2.5}$ or as an effect of an alternative co-pollutant. However, a commonly statistical issue in air pollution epidemiologic studies include measurement error across co-pollutants and correlation and multi-collinearity among co-pollutants.[34] That the research community has yet to reach consensus on gold-standard modeling and choice of whether to adjust for co-pollutants, suggests further research is needed. In terms of the current findings, it also suggests this to be a potential limitation. Additional prominent statistical issues include time windows in the concentration-response function, confounding, and spatial analysis.[34]

Additionally, the GEOS-CHEM method of quantifying $PM_{2.5}$ constituents does not include trace metals,[35] which are equally as important constituents of $PM_{2.5}$ that could add another layer to our understanding. $PM_{2.5}$ constituents that contribute very little to the total $PM_{2.5}$ mass (such as transition metals) are plagued by instrument-related measurement error; due to their low concentrations, methods of measuring them may be insensitive to such a small concentration, and will not demonstrate their true association with adverse health outcomes.[6]

Collinearity amongst constituents provides a serious limitation towards the measurement of individual constituent impact on health. For instance, an observed effect may represent the effect of another constituent that covaries with the measured constituent, or the effect of a combination of constituents that interact with a covarying constituent.[16] For instance, acids tend to be highly correlated with $PM_{2.5}$ mass, leading to multicollinearity that results in highly unstable coefficients.[18] High correlation amongst $PM_{2.5}$ constituents generally leads to overestimation of the contribution of each constituent when regressed separately in the model.[6,36] As a result, a non-toxic constituent may appear to be associated with adverse health effects, when an unobserved but correlated constituent is truly responsible for the toxic effect. Previous research has suggested that $PM_{2.5}$ constituents may be associated with health effects only because they are so highly correlated with $PM_{2.5}$ mass.[6]

Measurement error for constituents has also been suggested to be greater compared to that for $PM_{2.5}$ mass, particularly as spatial variability may differ amongst constituents.[27] The degree of measurement error may also differ between constituents, leading to wrongly stronger coefficients of some constituents because of having less measurement error, rather than an actual stronger effect.[16] Furthermore, modeling errors, due to differences in the understanding of the chemical processes between constituents, can also serve as a major issue as different CTMs, although perform similarly for estimating $PM_{2.5}$, performed variably for constituents.[27]

5.6 Conclusion

The present study provides no evidence that any particular constituent of $PM_{2.5}$ is more strongly associated with incidence of T2DM. Given concerns with best suitable statistical methods and exposure assignments, results are not confidently conclusive. More complex models and consensus on best exposure assignment may be needed to properly characterize which constituents of $PM_{2.5}$ are largely responsible for the adverse effects of $PM_{2.5}$ on T2DM development.

References for Article 3

1. World Health Organization. Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization; 2006.

2. van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC, et al. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environ Sci Technol. 2016 Apr 5;50(7):3762–72.

3. Brook RD, Newby DE, Rajagopalan S. Air Pollution and Cardiometabolic Disease: An Update and Call for Clinical Trials. Am J Hypertens. 2017 Dec 8;31(1):1–10.

4. National Research Council (NRC). Research Priorities for Airborne Particulate Matter: IV. Continuing Research Progress [Internet]. 2004 [cited 2020 May 11]. Available from: https://www.nap.edu/catalog/10957/research-priorities-for-airborne-particulate-matter-iv-continuing-research-progress

5. Lippmann M, Chen L-C, Gordon T, Ito K, Thurston GD. National Particle Component Toxicity (NPACT) Initiative: integrated epidemiologic and toxicologic studies of the health effects of particulate matter components. Res Rep Health Eff Inst. 2013 Oct;(177):5–13.

6. Krall JR, Chang HH, Sarnat SE, Peng RD, Waller LA. Current Methods and Challenges for Epidemiological Studies of the Associations Between Chemical Constituents of Particulate Matter and Health. Curr Environ Health Rep. 2015 Dec;2(4):388–98.

7. Chen H, Zhang Z, van Donkelaar A, Bai L, Martin RV, Lavigne E, et al. Understanding the joint impacts of fine particulate matter concentration and composition on the incidence and mortality of cardiovascular disease: a component-adjusted approach. Environ Sci Technol. 2020 Feb 26;

8. Geng G, Xiao Q, Zheng Y, Tong D, Zhang Y, Zhang X, et al. Impact of China's Air Pollution Prevention and Control Action Plan on PM2.5 chemical composition over eastern China. Sci China Earth Sci. 2019;

9. Wyzga RE, Rohr AC. Long-term particulate matter exposure: Attributing health effects to individual PM components. J Air Waste Manag Assoc. 2015 May;65(5):523–43.

10. Fu C, Chen Y, Wang F, Wang X, Song J, Jiang Q. High prevalence of hyperglycaemia and the impact of high household income in transforming rural China. BMC Public Health. 2011 Nov 12;11:862.

11. Boys BL, Martin RV, van Donkelaar A, MacDonell RJ, Hsu NC, Cooper MJ, et al. Fifteen-Year Global Time Series of Satellite-Derived Fine Particulate Matter. Environ Sci Technol. 2014 Oct 7;48(19):11109–18.

12. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. Environ Health Perspect. 2015 Feb;123(2):135–43.

13. GBD Maps Working Group. Burden of Disease Attributable to Coal-Burning and Other Air Pollution Sources in China. Boston, MA: Health Effects Institute; 2016. Report No.: 20.

14. Ma Z, Hu X, Sayer AM, Levy R, Zhang Q, Xue Y, et al. Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004–2013. Environ Health Perspect. 2016 Feb;124(2):184–92.

15. Liu C, Cai J, Qiao L, Wang H, Xu W, Li H, et al. The Acute Effects of Fine Particulate Matter Constituents on Blood Inflammation and Coagulation. Environ Sci Technol. 2017 Jul 18;51(14):8128–37.

16. Mostofsky E, Schwartz J, Coull BA, Koutrakis P, Wellenius GA, Suh HH, et al. Modeling the association between particle constituents of air pollution and health outcomes. Am J Epidemiol. 2012 Aug 15;176(4):317–26.

17. Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd ed. Springer International Publishing; 2015. (Springer Series in Statistics).

18. Thomas D. Statistical Methods in Environmental Epidemiology. New York, USA: Oxford University Press; 2009. 352 p.

19. Motulsky H. Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking. New York, NY: Oxford University Press; 1995.

20. Martinussen T, Scheike TH. Dynamic Regression Models for Survival Data. New York: Springer-Verlag; 2006. (Statistics for Biology and Health).

21. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag; 2000. (Statistics for Biology and Health).

22. Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. Ambient PM2.5, O_3, and NO_2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ Health Perspect. 2015 Nov;123(11):1180–6.

23. Yang Y, Ruan Z, Wang X, Yang Y, Mason TG, Lin H, et al. Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis. Environ Pollut. 2019 Apr;247:874–82.

24. Achilleos S, Kioumourtzoglou M-A, Wu C-D, Schwartz JD, Koutrakis P, Papatheodorou SI. Acute effects of fine particulate matter constituents on mortality: a systematic review and meta-regression analysis. Environ Int. 2017 Dec;109:89–100.

25. Atkinson RW, Mills IC, Walton HA, Anderson HR. Fine particle components and health--a systematic review and meta-analysis of epidemiological time series studies of daily mortality and hospital admissions. J Expo Sci Environ Epidemiol. 2015 Apr;25(2):208–14.

26. Lu F, Xu D, Cheng Y, Dong S, Guo C, Jiang X, et al. Systematic review and meta-analysis of the adverse health effects of ambient PM2.5 and PM10 pollution in the Chinese population. Environ Res. 2015 Jan;136:196–204.

27. US EPA, Sacks J. Integrated Science Assessment (ISA) for Particulate Matter (Final Report, 2019). Washington, DC: US Environmental Protection Agency; 2019. Report No.: EPA/600/R-19/188.

28. Cassee FR, Héroux M-E, Gerlofs-Nijland ME, Kelly FJ. Particulate matter beyond mass: recent health evidence on the role of fractions, chemical constituents and sources of emission. Inhal Toxicol. 2013 Dec;25(14):802–12.

29. Johnson TM, Guttikunda S, Wells GJ, Artaxo P, Bond TC, Russell AG, et al. Tools for improving air quality management : a review of top-down source apportionment techniques and their application in developing countries. The World Bank; 2011 Mar p. 1–220. Report No.: 81740.

30. Zhu Y, Huang L, Li J, Ying Q, Zhang H, Liu X, et al. Sources of particulate matter in China: Insights from source apportionment studies published in 1987–2017. Environment International. 2018 Jun 1;115:343–57.

31. Cao G, Zhang X, Zheng F. Inventory of black carbon and organic carbon emissions from China. Atmospheric Environment. 2006 Nov 1;40(34):6516–27.

32. Health Effects Institute. State of Global Air 2019: A Special Report on Global Exposure to Air Pollution and Its Disease Burden. Boston, MA: Health Effects Institute; 2019.

33. Zhang Y-L, Cao F. Is it time to tackle PM2.5 air pollutions in China from biomass-burning emissions? Environ Pollut. 2015 Jul 1;202:217–9.

34. Bateson TF, Coull BA, Hubbell B, Ito K, Jerrett M, Lumley T, et al. Panel discussion review: session three--issues involved in interpretation of epidemiologic analyses--statistical modeling. J Expo Sci Environ Epidemiol. 2007 Dec;17 Suppl 2:S90-96.

35. Yang Y, Pun VC, Sun S, Lin H, Mason TG, Qiu H. Particulate matter components and health: a literature review on exposure assessment. J Public Health Emerg [Internet]. 2018 Jan 4 [cited 2020 May 7];2(3). Available from: http://jphe.amegroups.com/article/view/4438

36. Lipfert FW, Wyzga RE. Statistical considerations in determining the health significance of constituents of airborne particulate matter. J Air Waste Manag Assoc. 1999;49(9):182–91.

Supplementary Material for Article 3

Supplementary Material 1. Constituent Modelling

Table S1. Three Options to Model the Association between Constituents and Outcomes: Advantages and Disadvantages

	Parameter and Model	Advantages	Disadvantages
1	**Constituent concentration** $g(\mu) = \beta 0 + \beta 1(\text{constituent}) + [\gamma' X]$		
	What is the effect of a particular constituent?	– Easy to interpret	– Confounding by total $PM_{2.5}$ – Confounding by other constituents that covary
2	**Constituent concentration adjusting for $PM_{2.5}$ mass** $g(\mu) = \beta 0 + \beta 1(\text{constituent}) + \beta 2(PM_{2.5}) + [\gamma' X]$		
	What is the effect of a particular constituent after adjusting for total $PM_{2.5}$?	– Accounts for total $PM_{2.5}$ – Easy to interpret	– Confounding by other constituents that co-vary – May be over adjusting for constituents highly correlated with total $PM_{2.5}$ – Collinearity between constituent and $PM_{2.5}$ and extent of this problem depend on relative contribution of constituent to total $PM_{2.5}$. – Does not provide indication of variation in constituent while holding total $PM_{2.5}$ constant.
3	**Constituent residual** $g(\mu) = \beta 0 + \beta 1(\text{residual}) + [\gamma' X]$ where residual: constituent = total $PM_{2.5}$		
	What is the effect of a particular constituent holding total $PM_{2.5}$ constant?	– Eliminates confounding by total $PM_{2.5}$ – No collinearity between constituent and $PM_{2.5}$ – Removes extraneous variation due to total $PM_{2.5}$.	– Hard to interpret: More of one constituent equates to less of another, which is larger problem for constituents with larger contribution to total $PM_{2.5}$. – Results do not provide information on absolute magnitude of constituent change associated with outcome.

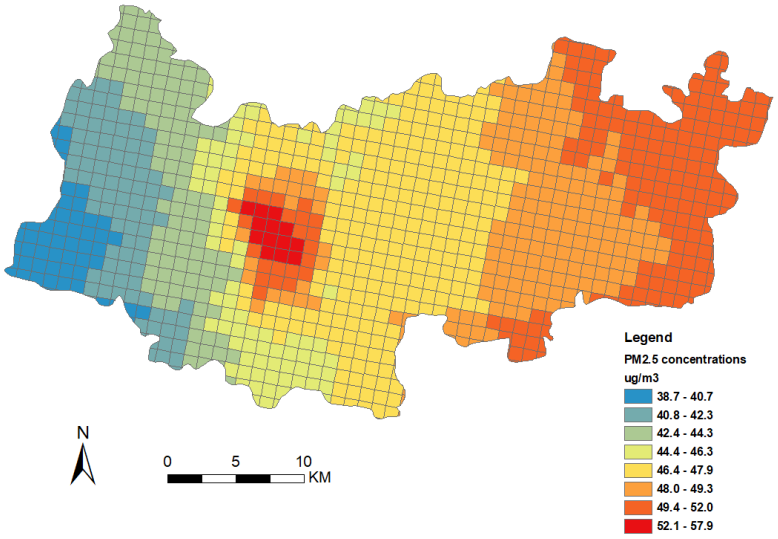
Modified from Mostofsky et al. (2012).[16]

Abbreviation: $PM_{2.5}$, ambient fine particulate matter.

Note: The models are represented in the form of a generalized linear model, with the dependent variable for the health outcome (μ) expressed as a function of the independent variables for constituents, total $PM_{2.5}$, and a matrix of [γ'X] other covariates. This model can accommodate different functional forms for the regression, such as linear regression, logistic regression, and survival analysis.

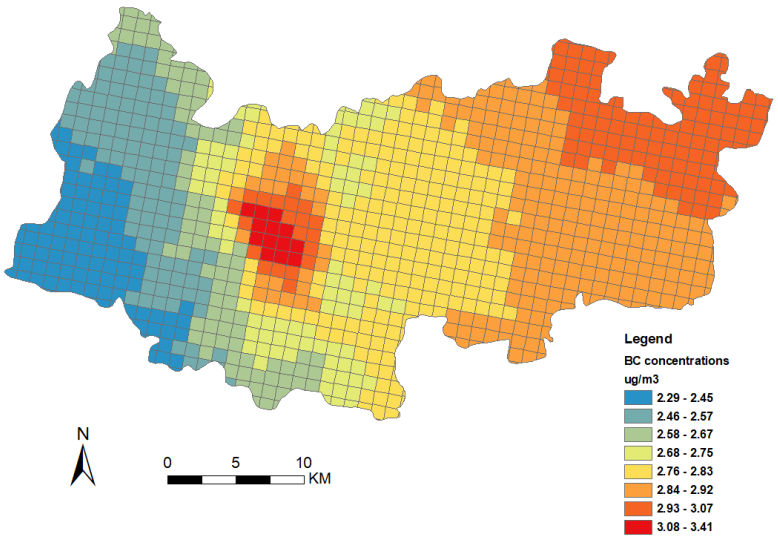
Supplementary Material 2. Spatial Distributions

Figure S2.1. Distribution of Average Annual Ambient Fine Particulate Matter Concentration over Region of Deqing County in 2006



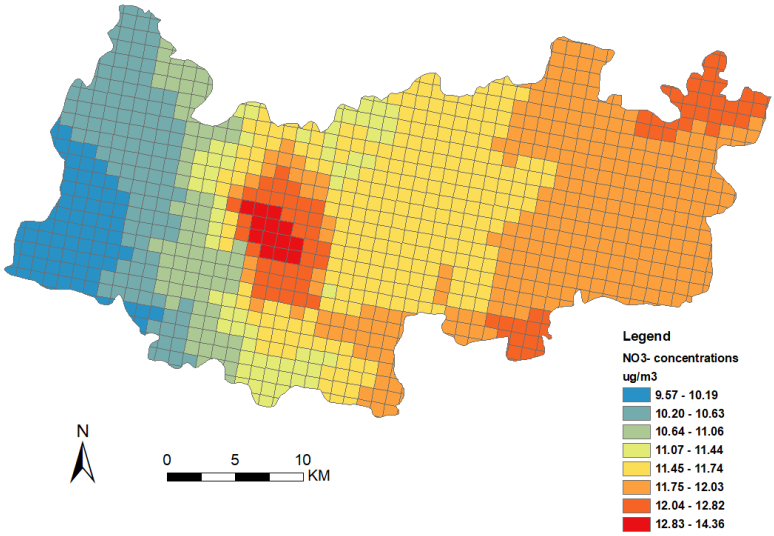
Abbreviation: $PM_{2.5}$, ambient fine particulate matter.

Figure S2.2. Distribution of Average Annual Black Carbon Concentration over Region of Deqing County in 2006



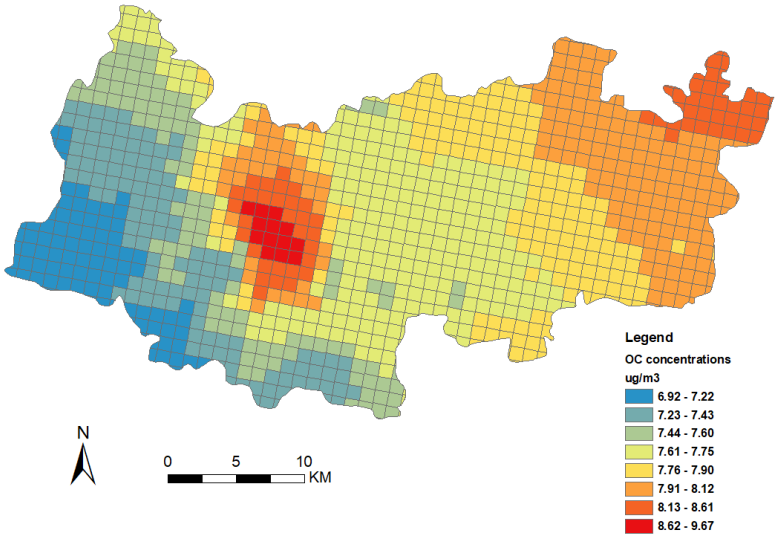
Abbreviation: BC, black carbon.

Figure S2.3. Distribution of Average Annual Nitrate Concentration over Region of Deqing County in 2006



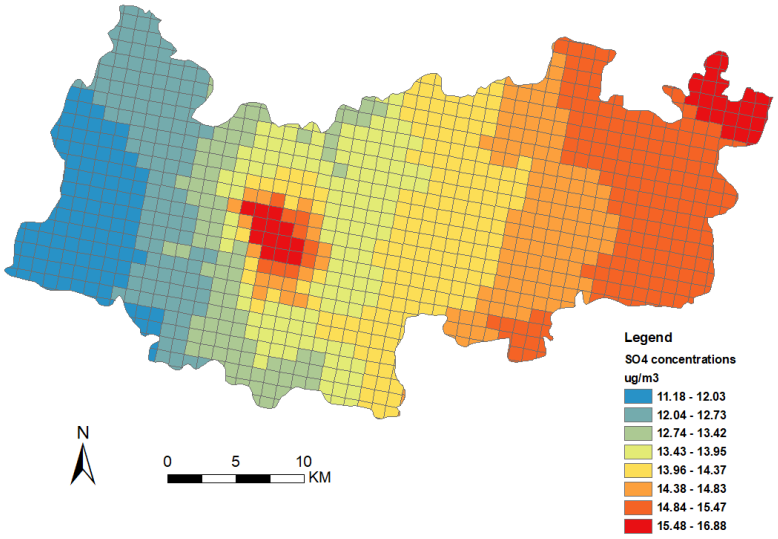
Abbreviations: NO_3^-, nitrate, measured in units of µg/m³.

Figure S2.4. Distribution of Average Annual Organic Carbon Concentration over Region of Deqing County in 2006



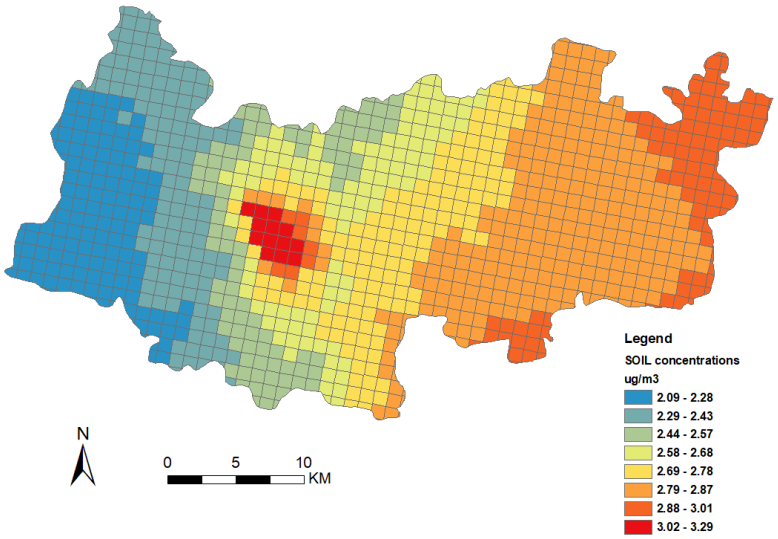
Abbreviations: OC, organic carbon, measured in units of µg/m³.

Figure S2.5. Distribution of Average Annual Sulfate Concentration over Region of Deqing County in 2006



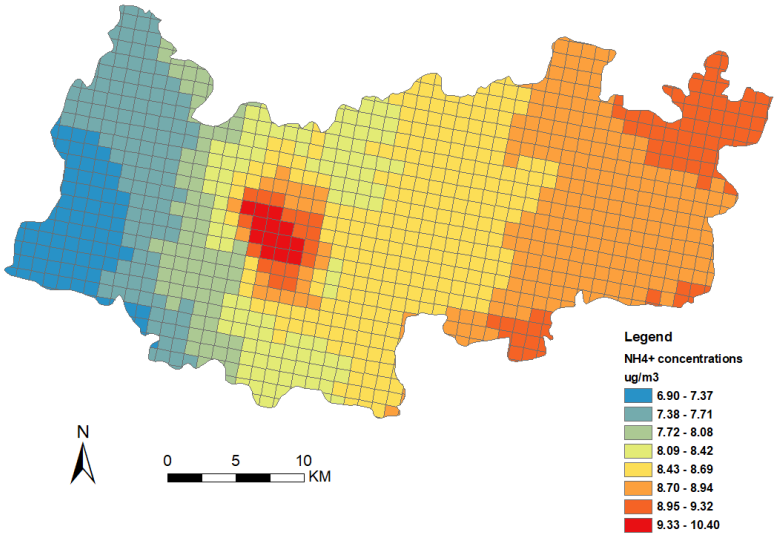
Abbreviations: SO_4^{2-}, sulfate, measured in units of µg/m³.

Figure S2.6. Distribution of Average Annual Soil Dust Concentration over Region of Deqing County in 2006



Soil dust is measured in units of µg/m³.

Figure S2.7. Distribution of Average Annual Ammonium Concentration over Region of Deqing County in 2006



Abbreviations: NH_4^+, ammonium, measured in units of $\mu g/m^3$.

Supplementary Material 3. Temporal Distributions

Table S3.1. Temporal Variability in Annual Ambient Fine Particulate Matter Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	52.61	3.53	43.86	50.94	53.29	55.17	62.54	4.22
By year								
2006	48.74	6.63	39.42	46.83	48.58	49.50	82.92	2.67
2007	61.41	5.34	42.17	59.08	61.92	63.17	84.00	4.08
2008	57.66	3.53	48.92	56.75	58.75	60.17	67.58	3.42
2009	52.16	3.77	42.67	49.25	53.00	55.67	61.17	6.42
2010	46.53	3.00	40.33	45.08	46.58	49.25	56.92	4.17
2011	49.34	3.27	41.33	47.00	49.67	52.50	60.08	5.50
2012	48.31	2.81	41.33	46.50	49.08	50.58	59.00	4.08
2013	53.76	4.58	41.50	51.42	55.17	57.83	63.25	6.42
2014	55.59	3.48	45.67	55.25	56.83	57.83	67.17	2.58

Abbreviations: SD, standard deviation; IQR, interquartile range.

Table S3.2. Temporal Variability in Annual Black Carbon Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	3.27	0.20	2.70	3.14	3.32	3.41	3.88	0.27
By year								
2006	2.85	0.31	2.03	2.77	2.86	2.89	4.35	0.13
2007	3.70	0.23	2.04	3.62	3.77	3.82	4.70	0.20
2008	3.54	0.21	2.98	3.44	3.65	3.67	4.14	0.23
2009	3.06	0.22	2.48	2.87	3.11	3.25	3.58	0.38
2010	2.88	0.18	2.48	2.76	2.90	3.03	3.51	0.27
2011	3.24	0.21	2.69	3.08	3.27	3.43	3.92	0.34
2012	3.18	0.18	2.67	3.07	3.26	3.32	3.87	0.25
2013	3.42	0.29	2.60	3.24	3.54	3.67	4.01	0.43
2014	3.53	0.22	2.87	3.48	3.62	3.66	4.27	0.17

Abbreviations: SD, standard deviation; IQR, interquartile range.

Table S3.3. Temporal Variability in Annual Nitrate Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	14.43	0.80	12.19	14.25	14.70	15.00	17.47	0.75
By year								
2006	11.66	0.74	6.31	11.57	11.79	11.87	14.43	0.31
2007	15.51	0.84	7.06	15.20	15.87	15.98	18.14	0.78
2008	15.82	0.89	13.49	15.73	16.14	16.32	18.80	0.58
2009	14.57	0.93	12.00	14.08	14.83	15.36	17.28	1.28
2010	12.95	0.77	11.26	12.76	13.01	13.58	16.04	0.82
2011	14.51	0.86	12.33	13.97	14.64	15.26	17.89	1.28
2012	14.36	0.76	12.48	13.84	14.48	14.84	17.76	1.00
2013	14.95	1.14	11.76	14.63	15.35	15.84	17.88	1.22
2014	15.53	0.94	12.88	15.45	15.88	15.99	19.10	0.54

Abbreviations: SD, standard deviation; IQR, interquartile range.

Table S3.4. Temporal Variability in Annual Organic Carbon Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	9.18	0.51	8.14	8.94	9.17	9.41	11.10	0.48
By year								
2006	8.11	1.62	7.01	7.68	7.82	7.93	17.12	0.25
2007	10.49	1.29	9.52	10.01	10.34	10.50	17.25	0.49
2008	10.44	0.41	9.46	10.19	10.58	10.69	12.39	0.50
2009	9.24	0.45	8.09	8.90	9.26	9.65	11.00	0.75
2010	7.70	0.32	7.12	7.50	7.67	7.98	9.53	0.47
2011	9.09	0.45	8.11	8.74	9.05	9.50	11.21	0.76
2012	8.33	0.33	7.56	8.13	8.39	8.53	10.27	0.41
2013	9.75	0.59	8.06	9.30	9.85	10.30	11.60	1.00
2014	9.46	0.42	8.22	9.32	9.57	9.72	11.62	0.41

Abbreviations: SD, standard deviation; IQR, interquartile range.

Table S3.5. Temporal Variability in Annual Sulfate Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	13.53	1.07	10.97	12.89	13.75	14.39	15.71	1.50
By year								
2006	14.37	1.47	11.38	13.84	14.57	15.02	20.26	1.18
2007	16.98	1.29	12.72	16.32	17.34	17.93	19.52	1.62
2008	15.02	1.17	12.36	14.54	15.52	15.86	17.15	1.32
2009	13.12	1.19	10.36	12.17	13.32	14.27	14.93	2.10
2010	11.00	0.88	9.20	10.49	11.07	11.83	13.13	1.33
2011	11.53	0.94	9.36	10.84	11.66	12.46	13.72	1.62
2012	12.00	0.88	9.94	11.42	12.23	12.73	14.29	1.32
2013	13.23	1.36	9.87	12.42	13.62	14.48	15.18	2.06
2014	14.49	1.05	11.59	14.30	14.88	15.21	17.17	0.91

Abbreviations: SD, standard deviation; IQR, interquartile range.

Table S3.6. Temporal Variability in Annual Soil Dust Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	3.11	0.60	2.44	2.84	3.10	3.20	6.21	0.36
By year								
2006	3.17	2.40	2.13	2.62	2.81	2.86	16.72	0.24
2007	3.95	2.35	3.02	3.33	3.60	3.68	16.81	0.35
2008	2.85	0.20	2.39	2.78	2.93	2.99	3.33	0.21
2009	3.24	0.28	2.55	3.04	3.33	3.46	3.86	0.42
2010	4.27	0.34	3.59	4.06	4.32	4.57	5.16	0.51
2011	2.59	0.25	2.04	2.37	2.65	2.82	3.13	0.45
2012	1.89	0.15	1.55	1.79	1.94	2.01	2.30	0.22
2013	3.13	0.32	2.30	2.97	3.26	3.40	3.65	0.43
2014	2.91	0.25	2.35	2.70	2.99	3.07	3.47	0.38

Abbreviations: SD, standard deviation; IQR, interquartile range.

Table S3.7. Temporal Variability in Annual Ammonium Concentration in Deqing County, from 2006 to 2014

	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
Overall	9.16	0.59	7.60	8.94	9.33	9.64	10.90	0.70
By year								
2006	8.64	0.73	6.39	8.44	8.74	8.88	11.42	0.44
2007	10.74	0.66	6.81	10.46	10.95	11.17	12.27	0.71
2008	10.13	0.66	8.49	9.97	10.36	10.57	11.82	0.60
2009	9.06	0.68	7.33	8.60	9.21	9.70	10.57	1.10
2010	7.80	0.52	6.69	7.60	7.84	8.25	9.55	0.65
2011	8.46	0.57	7.08	8.10	8.55	9.02	10.28	0.92
2012	8.58	0.52	7.31	8.26	8.73	8.98	10.46	0.73
2013	9.20	0.80	7.06	8.84	9.44	9.89	10.81	1.05
2014	9.86	0.64	8.04	9.83	10.10	10.23	11.93	0.40

Abbreviations: SD, standard deviation; IQR, interquartile range.

Supplementary Material 4. Three-Year Average $PM_{2.5}$ and Its Constituents

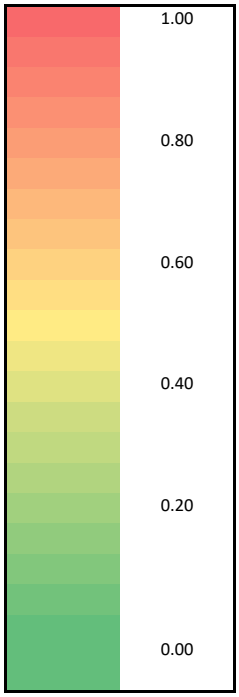
Table S4.1. Descriptive Statistics for Three-Year Average Concentration of Ambient Fine Particulate Matter and Its Constituents in Deqing County

Three-year average concentration ($\mu g/m^3$)	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
$PM_{2.5}$	52.82	3.69	40.86	50.94	53.22	55.67	75.61	4.72
BC	3.33	0.22	2.54	3.23	3.41	3.51	4.11	0.28
NO_3^-	14.76	0.94	9.79	14.48	14.93	15.48	18.25	1.00
OC	9.3	0.49	7.74	9.01	9.4	9.57	15.03	0.56
SO_4^{2-}	13.49	1.34	9.77	12.86	13.32	14.34	18.57	1.48
Soil dust	2.75	0.33	1.97	2.58	2.78	2.95	12.17	0.38
NH_4^+	9.25	0.67	7.13	8.99	9.28	9.77	11.5	0.79

Abbreviations: SD, standard deviation; IQR, interquartile range; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Figure S4.1. Correlation between Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

	$PM_{2.5}$	BC	NO_3^-	OC	SO_4^{2-}	Soil dust	NH_4^+
$PM_{2.5}$	1	0.94	0.83	0.95	0.93	0.82	0.99
BC	0.94	1	0.93	0.89	0.77	0.68	0.92
NO_3^-	0.83	0.93	1	0.78	0.57	0.53	0.81
OC	0.95	0.89	0.78	1	0.86	0.84	0.93
SO_4^{2-}	0.93	0.77	0.57	0.86	1	0.81	0.95
Soil dust	0.82	0.68	0.53	0.84	0.81	1	0.79
NH_4^+	0.99	0.92	0.81	0.93	0.95	0.79	1



Abbreviations: $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Table S4.2. Unadjusted Single-Constituent Models Using Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR	HR per IQR-increase	95% CI	p-value
1	$PM_{2.5}$	4.72	1.12	0.99, 1.27	0.0747
2	BC	0.28	1.20	1.06, 1.37	0.0055
3	NO_3^-	1.00	1.20	1.07, 1.33	0.0013
4	OC	0.56	1.06	0.95, 1.18	0.3098
5	SO_4^{2-}	1.48	1.02	0.92, 1.14	0.6482
6	Soil dust	0.38	1.09	1.02, 1.17	0.0126
7	NH_4^+	0.79	1.09	0.97, 1.21	0.1565

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase in exposure**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Table S4.3. Parsimoniously-adjusted Single-Constituent Models Using Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR	HR per IQR-increase	95% CI	p-value
1	$PM_{2.5}$	4.72	1.20	1.06, 1.38	0.0059
2	BC	0.28	1.22	1.07, 1.39	0.0035
3	NO_3^-	1.00	1.17	1.05, 1.32	0.0059
4	OC	0.56	1.11	1.00, 1.24	0.0588
5	SO_4^{2-}	1.48	1.16	1.02, 1.31	0.0200
6	Soil dust	0.38	1.12	1.06, 1.18	<.0001
7	NH_4^+	0.79	1.17	1.03, 1.32	0.0130

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase in exposure**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, passive smoking exposure, vegetable consumption, history of disease, and occupation.

Table S4.4. Unadjusted Constituent-$PM_{2.5}$ Joint Models Using Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC + $PM_{2.5}$	0.28	1.99	1.34, 2.95	0.0007
2	NO_3^-+ $PM_{2.5}$	1.00	1.41	1.14, 1.74	0.0014
3	OC + $PM_{2.5}$	0.56	0.71	0.50, 1.01	0.0568
4	SO_4^{2-} + $PM_{2.5}$	1.48	0.61	0.46, 0.82	0.0009
5	Soil dust + $PM_{2.5}$	0.38	1.07	0.97, 1.17	0.1815
6	NH_4^++ $PM_{2.5}$	0.79	0.52	0.32, 0.83	0.0094

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Table S4.5. Parsimoniously-adjusted Constituent-$PM_{2.5}$ Joint Models Using Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC + $PM_{2.5}$	0.28	1.41	0.78, 2.55	0.2522
2	NO_3^-+ $PM_{2.5}$	1.00	1.09	0.81, 1.46	5772
3	OC + $PM_{2.5}$	0.56	0.68	0.48, 0.98	0.0384
4	SO_4^{2-} + $PM_{2.5}$	1.48	0.82	0.56, 1.22	0.3291
5	Soil dust + $PM_{2.5}$	0.38	1.09	1.00, 1.18	0.0385
6	NH_4^++ $PM_{2.5}$	0.79	0.62	0.35, 1.09	0.0979

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, passive smoking exposure, vegetable consumption, history of disease, and occupation.

Table S4.6. Unadjusted Constituent Residual Model Using Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC residual	0.28	0.97	0.95, 0.99	0.0146
2	NO_3^- residual	1.00	0.96	0.91, 1.01	0.1192
3	OC residual	0.56	1.06	1.01, 1.11	0.0126
4	SO_4^{2-} residual	1.48	1.22	1.10, 1.36	0.0001
5	Soil dust residual	0.38	1.00	0.98, 1.01	0.6500
6	NH_4^+ residual	0.79	1.15	1.05, 1.26	0.0025

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Table S4.7. Parsimoniously-adjusted Constituent Residual Model Using Three-Year Average-Assigned Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC residual	0.28	1.00	0.97, 1.04	0.9847
2	NO_3^- residual	1.00	1.05	0.98, 1.12	0.1629
3	OC residual	0.56	1.08	1.02, 1.13	0.0033
4	SO_4^{2-} residual	1.48	1.17	1.03, 1.32	0.0134
5	Soil dust residual	0.38	0.99	0.98, 1.01	0.3562
6	NH_4^+ residual	0.79	1.14	1.03, 1.26	0.0155

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, passive smoking exposure, vegetable consumption, history of disease, and occupation.

Supplementary Material 5. Yearly Time-Varying $PM_{2.5}$ and Its Constituents

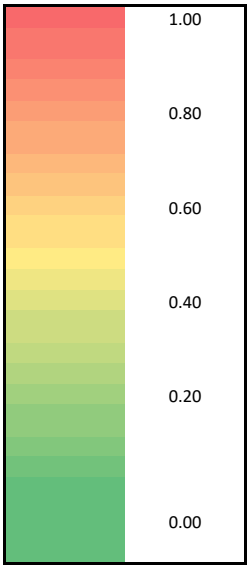
Table S5.1. Descriptive Statistics for Time-Varying Concentration of Ambient Fine Particulate Matter and Its Constituents in Deqing County

Time-varying concentration ($\mu g/m^3$)	Mean	SD	Minimum	Quartiles			Maximum	IQR
				Lower	Median	Upper		
$PM_{2.5}$	51.80	5.27	39.42	47.58	52.08	56.00	84.00	8.42
BC	3.27	0.33	2.03	3.03	3.29	3.57	4.35	0.53
NO_3^-	14.62	1.35	6.31	13.73	14.96	15.65	19.10	1.92
OC	9.19	0.83	7.01	8.58	9.38	9.72	17.25	1.14
SO_4^{2-}	13.02	1.81	9.20	11.62	12.96	14.54	20.26	2.93
Soil dust	2.75	0.60	1.55	2.39	2.78	3.13	16.81	0.74
NH_4^+	9.04	0.97	6.39	8.38	9.05	9.88	12.27	1.51

Abbreviations: SD, standard deviation; IQR, interquartile range; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Figure S5.1. Correlation between Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

	$PM_{2.5}$	BC	NO_3^-	OC	SO_4^{2-}	Soil dust	NH_4^+
$PM_{2.5}$	1	0.96	0.89	0.88	0.93	0.48	0.99
BC	0.96	1	0.93	0.85	0.85	0.35	0.95
NO_3^-	0.89	0.93	1	0.84	0.71	0.23	0.88
OC	0.88	0.85	0.84	1	0.73	0.35	0.83
SO_4^{2-}	0.93	0.85	0.71	0.73	1	0.43	0.96
Soil dust	0.48	0.35	0.23	0.35	0.43	1	0.38
NH_4^+	0.99	0.95	0.88	0.83	0.96	0.38	1



Abbreviations: $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.

Table S5.2. Unadjusted Single-Constituent Models Using Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR	HR per IQR-increase	95% CI	p-value
1	$PM_{2.5}$	8.42	1.45	1.16, 1.80	0.0011
2	BC	0.53	1.46	1.17, 1.83	0.0009
3	NO_3^-	1.92	1.37	1.13, 1.66	0.0014
4	OC	1.14	1.25	1.01, 1.56	0.0409
5	SO_4^{2-}	2.93	1.61	1.24, 2.10	0.0003
6	Soil dust	0.74	1.19	1.07, 1.32	0.0017
7	NH_4^+	1.51	1.47	1.18, 1.84	0.0007

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Table S5.3. Parsimoniously-adjusted Single-Constituent Models Using Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR	HR per IQR-increase	95% CI	p-value
1	$PM_{2.5}$	8.42	1.39	1.11, 1.75	0.0049
2	BC	0.53	1.41	1.12, 1.79	0.0037
3	NO_3^-	1.92	1.33	1.09, 1.63	0.0051
4	OC	1.14	1.23	0.98, 1.54	0.0751
5	SO_4^{2-}	2.93	1.53	1.17, 2.00	0.0021
6	Soil dust	0.74	1.17	1.04, 1.31	0.0089
7	NH_4^+	1.51	1.42	1.12, 1.79	0.0034

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, education, and occupation.

Table S5.4. Unadjusted Constituent-$PM_{2.5}$ Joint Models Using Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC + $PM_{2.5}$	0.53	1.93	0.30, 12.51	0.4897
2	NO_3^-+ $PM_{2.5}$	1.92	1.00	0.42, 2.40	0.9967
3	OC + $PM_{2.5}$	1.14	0.21	0.09, 0.49	0.0004
4	SO_4^{2-} + $PM_{2.5}$	2.93	8.88	1.56, 50.60	0.0139
5	Soil dust + $PM_{2.5}$	0.74	1.03	0.78, 1.35	0.8460
6	NH_4^++ $PM_{2.5}$	1.51	26.61	0.98, 724.24	0.0516

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Table S5.5. Parsimoniously-adjusted Constituent-$PM_{2.5}$ Joint Models Using Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC + $PM_{2.5}$	0.53	2.54	0.39, 16.74	0.3315
2	NO_3^-+ $PM_{2.5}$	1.92	1.10	0.44, 2.79	0.8384
3	OC + $PM_{2.5}$	1.14	0.26	0.11, 0.62	0.0025
4	SO_4^{2-} + $PM_{2.5}$	2.93	6.76	0.98, 33.99	0.0532
5	Soil dust + $PM_{2.5}$	0.74	1.02	0.77, 1.35	0.8939
6	NH_4^++ $PM_{2.5}$	1.51	15.86	0.56, 453.381	0.1062

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, education, and occupation.

Table S5.6. Unadjusted Constituent Residual Model Using Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC residual	0.53	1.04	0.94, 1.15	0.4821
2	NO_3^- residual	1.92	1.18	1.00, 1.40	0.0567
3	OC residual	1.14	1.23	1.12, 1.34	<.0001
4	SO_4^{2-} residual	2.93	1.24	0.91, 1.69	0.1770
5	Soil dust residual	0.74	1.04	1.02, 1.07	0.0021
6	NH_4^+ residual	1.51	1.03	0.67, 1.53	0.9023

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase in exposure**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*

Table S5.7. Parsimoniously-adjusted Constituent Residual Model Using Time-Varying Exposure to Ambient Fine Particulate Matter and Its Constituents

Model	Variables in model	IQR of constituent	HR per IQR-increase	95% CI	p-value
1	BC residual	0.53	1.01	0.91, 1.13	0.8477
2	NO_3^- residual	1.92	1.15	0.95, 1.38	0.1550
3	OC residual	1.14	1.20	1.10, 1.31	<.0001
4	SO_4^{2-} residual	2.93	1.23	0.89, 1.70	0.2075
5	Soil dust residual	0.74	1.04	1.01, 1.07	0.0082
6	NH_4^+ residual	1.51	1.03	0.69, 1.52	0.8945

*Abbreviations: IQR, interquartile range; HR, hazard ratio, **per IQR-increase in exposure**; CI, confidence interval; $PM_{2.5}$, ambient fine particulate matter; BC, black carbon; NO_3^-, nitrate; OC, organic carbon; SO_4^{2-}, sulfate; NH_4^+, ammonium.*
Covariates in the parsimoniously-adjusted model include age, education, and occupation.

Supplementary Material 6. Proportional Hazards Assumption for Constituents

Figure S6.1. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Ambient Fine Particulate Matter

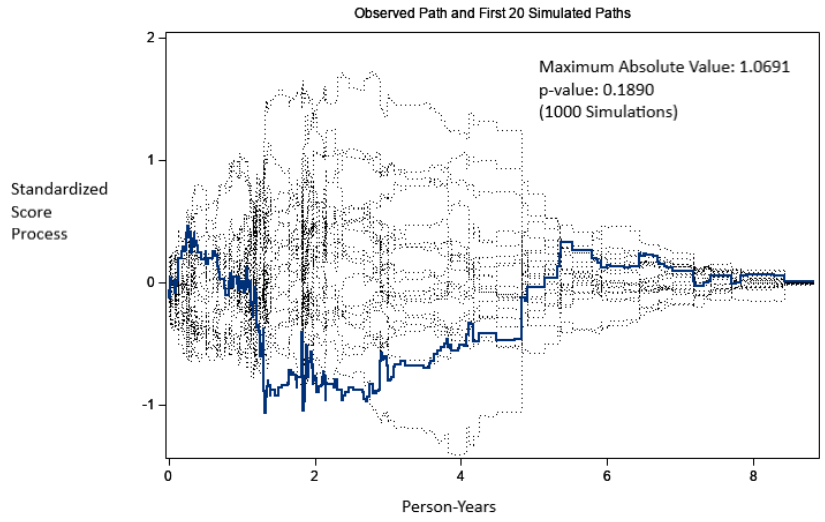


Figure S6.2. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Black Carbon

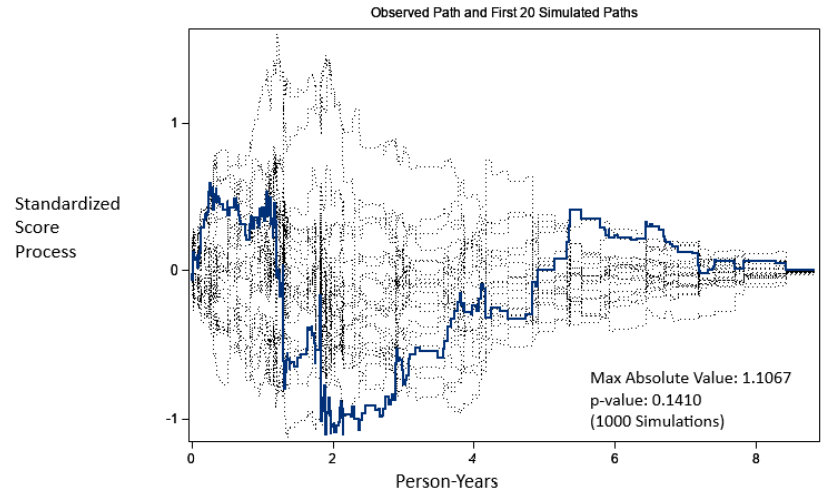


Figure S6.3. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Nitrate

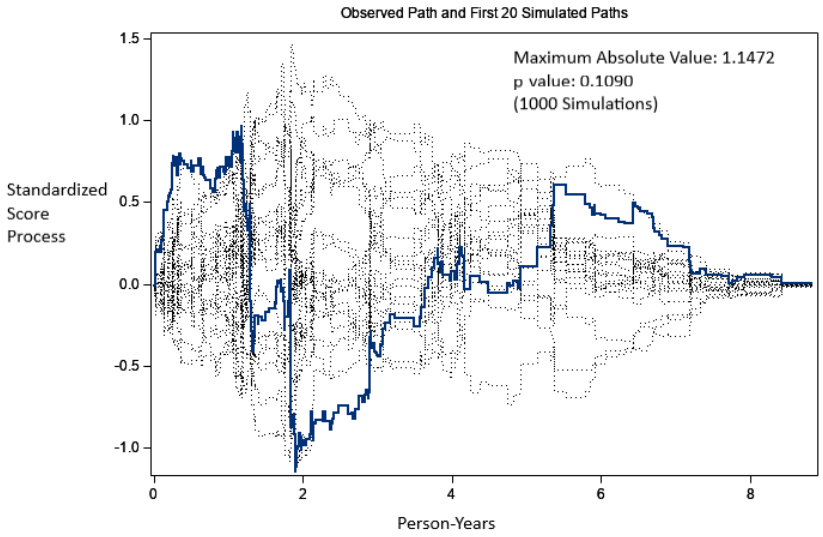


Figure S6.4. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Organic Carbon

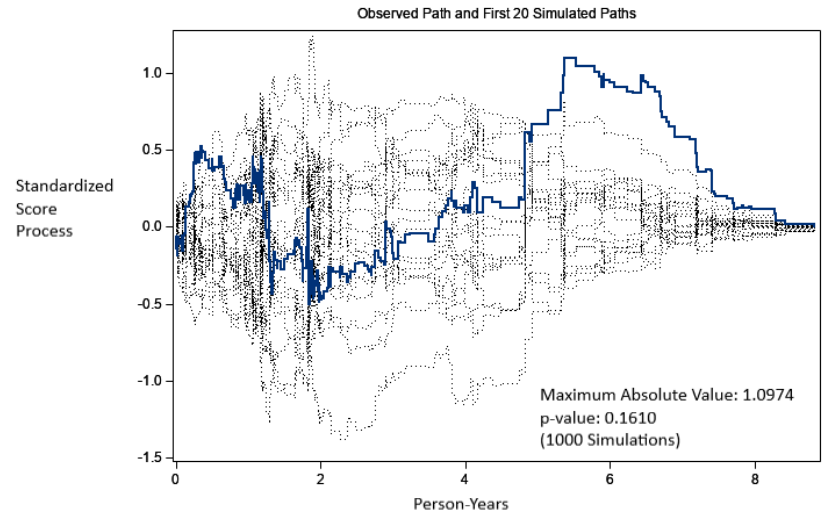


Figure S6.5. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Sulfate

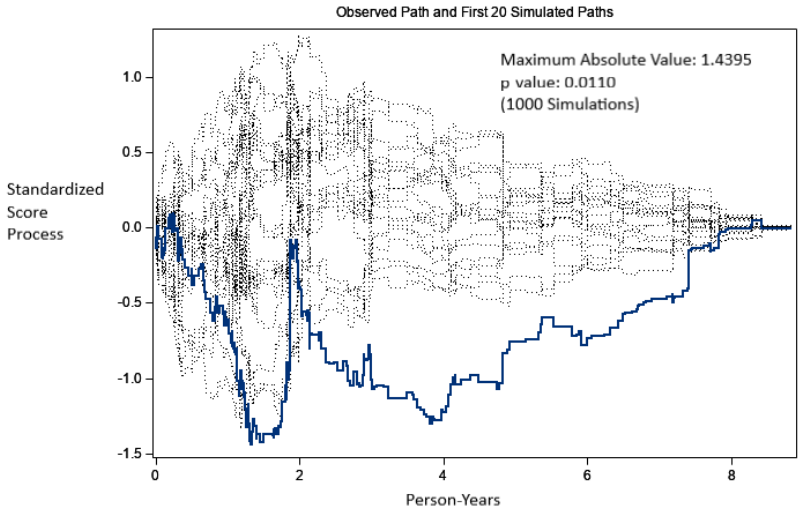


Figure S6.6. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Soil Dust

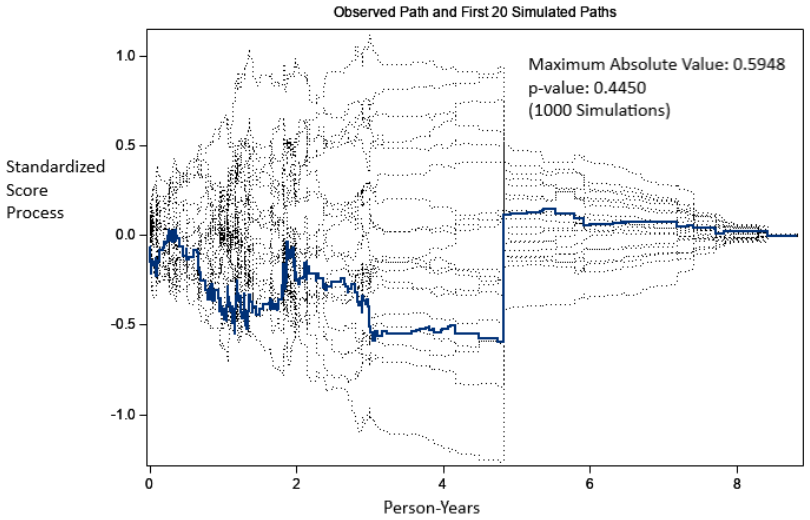
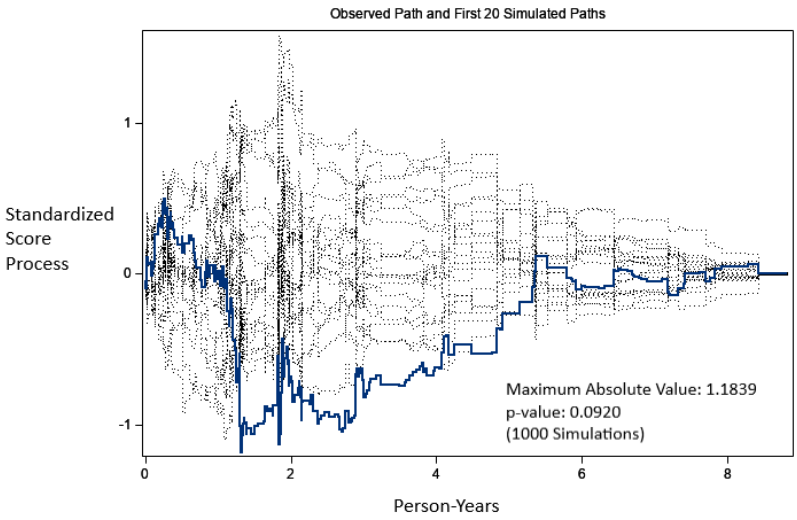


Figure S6.7. <ASSESS> Statement for Type 2 Diabetes Mellitus Associated with Baseline Exposure to Ammonium



_____ End of Article 3 _____

CHAPTER 6. OVERALL CONCLUSIONS

The prevalence of T2DM is on the rise worldwide, with major spikes in China in particular.[2] Results of a nationally representative epidemiological survey suggested that the prevalence of T2DM in 2017 was anywhere from 11.2% to 12.8%, depending on whether WHO or ADA diagnostic criteria was used.[27] Attention towards long-term exposure to $PM_{2.5}$ as a risk factor for developing T2DM began as researchers realized that the major risk factors commonly associated with T2DM alone may not be fully responsible for the rise in the risk of T2DM worldwide. The 2017 Lancet Commission report on pollution and health urged researchers to question and seek answers on the causal links between air pollution and T2DM, particularly in low- and middle-income countries.[28]

Epidemiologic research has suggested that long-term exposure to $PM_{2.5}$ is associated with an increased risk of developing T2DM; however, the majority of the studies are based on data from North American or European populations, which are exposure to differing sources and concentrations of $PM_{2.5}$ than their Asian counterparts,[24] as well as differing risk profiles for diabetes.[97,98] Furthermore, few studies have examined the association in rural populations in China, where exposure to air pollution remains elevated and should not be dismissed.[49] Our study of adult residents of rural Deqing County, Zhejiang, China, demonstrated that long-term exposure to $PM_{2.5}$ was positively associated with the incidence of T2DM; the adjusted HR for incident T2DM was 1.26 (95% CI: 1.02, 1.54) for every 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure (**Article 1**). The association between $PM_{2.5}$ and T2DM incidence was stronger in smokers (adjusted HR: 1.68, 95% CI: 1.12, 2.52) than in non-smokers (adjusted HR: 1.16, 95% CI: 0.91, 1.47).

Whether cigarette smoking worked jointly with $PM_{2.5}$ to lead to an increased risk of T2DM via a biological interaction was also assessed (**Article 2**). Effect modification is commonly assessed using multiplicative interactive terms in epidemiologic research.[99] However, it has been suggested that an

additive interaction term model better mimics a biologic interaction.[100] Our data did not provide evidence for a significant synergistic effect of cigarette smoking and $PM_{2.5}$ on the incidence of T2DM.

Finally, little is known about why $PM_{2.5}$ leads to increased risk of T2DM. Although $PM_{2.5}$ is commonly represented as total mass in epidemiologic studies, it has been suggested that $PM_{2.5}$ mass itself may not be responsible for the adverse health effects; rather, specific constituents of $PM_{2.5}$ may be more toxic and are what lead to the development of disease.[55] Whether specific constituents of $PM_{2.5}$ pose stronger links to the development of T2DM was examined (**Article 3**). Overall the results were inconclusive, largely expected to be due to issues of collinearity amongst $PM_{2.5}$ and its constituents. Other researchers have similarly concluded that although many constituents may contribute to poor health, the nature of the current evidence limits our ability to identify the specific constituent(s).[101]

The results of our study imply that the risk of T2DM related to increased exposure to $PM_{2.5}$ is just as much of a problem in rural regions of China as it is in urban China and developed high-income countries, where the majority of research on this topic has be published. Further research on a broader range of populations is required to fully characterize the risk of T2DM related to increased long-term exposure to $PM_{2.5.}$

REFERENCES

1. Diabetes Canada Clinical Practice Guidelines Expert Committee, Punthakee Z, Goldenberg R, Katz P. Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. Can J Diabetes. 2018 Apr;42 Suppl 1:S10–5.
2. World Health Organization. Global report on diabetes. World Health Organization; 2016.
3. Rajagopalan S, Brook RD. Air pollution and type 2 diabetes: mechanistic insights. Diabetes. 2012 Dec;61(12):3037–45.
4. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med. 1998 Jul;15(7):539–53.
5. WHO. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation. Geneva: World Health Organization; 2011. (WHO Guidelines Approved by the Guidelines Review Committee).
6. ADA. Standards of medical care in diabetes--2010. Diabetes Care. 2010 Jan;33 Suppl 1:S11-61.
7. ADA. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. Diabetes Care. 2019 Jan;42(Suppl 1):S13–28.
8. Cox ME, Edelman D. Tests for Screening and Diagnosis of Type 2 Diabetes. Clinical Diabetes. 2009 Oct 2;27(4):132–8.
9. Sacks DB. A1C versus glucose testing: a comparison. Diabetes Care. 2011 Feb;34(2):518–23.
10. Jia W, Weng J, Zhu D, Ji L, Lu J, Zhou Z, et al. Standards of medical care for type 2 diabetes in China 2019. Diabetes Metab Res Rev. 2019;35(6):e3158.
11. Maddaloni E, Pozzilli P. Why China guidelines for type 2 diabetes represent an opportunity for treating this disease. Diabetes Metab Res Rev. 2016;32(5):438–9.
12. Ma X, Zhang Y-L, Ji Q, Xing Y, Pan H, Chen S, et al. Diagnostic criteria for diabetes in China: are we pushing too much beyond evidence? Eur J Clin Nutr. 2017;71(7):812–5.
13. Weng J, Ji L, Jia W, Lu J, Zhou Z, Zou D, et al. Standards of care for type 2 diabetes in China. Diabetes Metab Res Rev. 2016;32(5):442–58.
14. Chan JCN, Chow EYK, Luk AOY. Diabetes in China and the Western Pacific Region. In: Dagogo-Jack S, editor. Diabetes Mellitus in Developing Countries and Underserved Communities. Cham: Springer International Publishing; 2017. p. 63–83.
15. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol. 2011 Nov 8;8(4):228–36.
16. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. PLoS ONE. 2018;13(3):e0194127.
17. Soares M, Müller MJ. Type 2 diabetes in Asia: where do we go from here? Eur J Clin Nutr. 2017;71(7):801–2.
18. Weber MB, Oza-Frank R, Staimez LR, Ali MK, Narayan KMV. Type 2 diabetes in Asians: prevalence, risk factors, and effectiveness of behavioral intervention at individual and population levels. Annu Rev Nutr. 2012 Aug 21;32:417–39.
19. Ma RCW. Epidemiology of diabetes and diabetic complications in China. Diabetologia. 2018;61(6):1249–60.

20. Chan JCN, Malik V, Jia W, Kadowaki T, Yajnik CS, Yoon K-H, et al. Diabetes in Asia: Epidemiology, Risk Factors, and Pathophysiology. JAMA. 2009 May 27;301(20):2129–40.

21. Zhang N, Du SM, Ma GS. Current lifestyle factors that increase risk of T2DM in China. Eur J Clin Nutr. 2017 Jul;71(7):832–8.

22. Yang G, Wang Y, Zeng Y, Gao GF, Liang X, Zhou M, et al. Rapid health transition in China, 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet. 2013 Jun 8;381(9882):1987–2015.

23. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018;14(2):88–98.

24. Thiering E, Heinrich J. Epidemiology of air pollution and diabetes. Trends Endocrinol Metab. 2015 Jul;26(7):384–94.

25. Chan M. China's Burgeoning Epidemic of Diabetes-Associated Mortality. JAMA. 2017 17;317(3):264–6.

26. Wang B, Xu D, Jing Z, Liu D, Yan S, Wang Y. Effect of long-term exposure to air pollution on type 2 diabetes mellitus risk: a systemic review and meta-analysis of cohort studies. Eur J Endocrinol. 2014 Nov;171(5):R173-182.

27. Li Y, Teng D, Shi X, Qin G, Qin Y, Quan H, et al. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. BMJ. 2020 28;369:m997.

28. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu NN, et al. The Lancet Commission on pollution and health. Lancet. 2018 03;391(10119):462–512.

29. Dendup T, Feng X, Clingan S, Astell-Burt T. Environmental Risk Factors for Developing Type 2 Diabetes Mellitus: A Systematic Review. Int J Environ Res Public Health. 2018 Jan 5;15(1).

30. Shen X, Vaidya A, Wu S, Gao X. THE DIABETES EPIDEMIC IN CHINA: AN INTEGRATED REVIEW OF NATIONAL SURVEYS. Endocr Pract. 2016 Sep;22(9):1119–29.

31. Health Effects Institute. State of Global Air 2019: A Special Report on Global Exposure to Air Pollution and Its Disease Burden. Boston, MA: Health Effects Institute; 2019.

32. Johnson TM, Guttikunda S, Wells GJ, Artaxo P, Bond TC, Russell AG, et al. Tools for improving air quality management : a review of top-down source apportionment techniques and their application in developing countries. The World Bank; 2011 Mar p. 1–220. Report No.: 81740.

33. Rajagopalan S, Al-Kindi SG, Brook RD. Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review. J Am Coll Cardiol. 2018 23;72(17):2054–70.

34. World Health Organization. Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization; 2006.

35. World Health Organization. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. Geneva, Switzerland: World Health Organization; 2006.

36. Huang G. PM2.5 opened a door to public participation addressing environmental challenges in China. Environ Pollut. 2015 Feb 1;197:313–5.

37. Yuan Y, Liu S, Castro R, Pan X. PM2.5 Monitoring and Mitigation in the Cities of China. Environ Sci Technol. 2012 Apr 3;46(7):3627–8.

38. Huang J, Pan X, Guo X, Li G. Health impact of China's Air Pollution Prevention and Control Action Plan: an analysis of national air quality monitoring and mortality data. Lancet Planet Health. 2018 Jul 1;2(7):e313–23.

39. Ma Z, Hu X, Sayer AM, Levy R, Zhang Q, Xue Y, et al. Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004–2013. Environ Health Perspect. 2016 Feb;124(2):184–92.

40. van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC, et al. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environ Sci Technol. 2016 Apr 5;50(7):3762–72.

41. Boys BL, Martin RV, van Donkelaar A, MacDonell RJ, Hsu NC, Cooper MJ, et al. Fifteen-Year Global Time Series of Satellite-Derived Fine Particulate Matter. Environ Sci Technol. 2014 Oct 7;48(19):11109–18.

42. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. Environ Health Perspect. 2015 Feb;123(2):135–43.

43. Philip S, Martin RV, van Donkelaar A, Lo JW-H, Wang Y, Chen D, et al. Global Chemical Composition of Ambient Fine Particulate Matter for Exposure Assessment. Environ Sci Technol. 2014 Nov 18;48(22):13060–8.

44. Vital Strategies. Accelerating City Progress on Clean Air: Innovation and Action Guide [Internet]. New York, NY: Vital Strategies; 2020 [cited 2020 Apr 30]. Available from: https://www.vitalstrategies.org/resources/accelerating-city-progress-on-clean-air-innovation-and-action-guide/

45. Bragg F, Holmes MV, Iona A, Guo Y, Du H, Chen Y, et al. Association Between Diabetes and Cause-Specific Mortality in Rural and Urban Areas of China. JAMA. 2017 17;317(3):280–9.

46. Jin C, Cheng J, Lu Y, Huang Z, Cao F. Spatial inequity in access to healthcare facilities at a county level in a developing country: a case study of Deqing County, Zhejiang, China. Int J Equity Health. 2015 Aug 19;14:67.

47. Feng R, Li K, Cheng J, Xie S, Chai J, Wei P, et al. Toward integrated and sustainable prevention against diabetes in rural China: study rationale and protocol of eCROPS. BMC Endocr Disord. 2013 Aug 7;13:28.

48. Liu X, Li Y, Li L, Zhang L, Ren Y, Zhou H, et al. Prevalence, awareness, treatment, control of type 2 diabetes mellitus and risk factors in Chinese rural population: the RuralDiab study. Sci Rep. 2016 11;6:31426.

49. Rafaj P, Kiesewetter G, Gül T, Schöpp W, Cofala J, Klimont Z, et al. Outlook for clean air in the context of sustainable development goals. Global Environmental Change. 2018 Nov 1;53:1–11.

50. Zhao X, Zhang S, Fan C. Environmental externality and inequality in China: current status and future choices. Environ Pollut. 2014 Jul;190:176–9.

51. Rohde RA, Muller RA. Air Pollution in China: Mapping of Concentrations and Sources. PLoS ONE. 2015;10(8):e0135749.

52. Davidson CI, Phalen RF, Solomon PA. Airborne Particulate Matter and Human Health: A Review. Aerosol Sci Tech. 2005 Aug 1;39(8):737–49.

53. Brown JS, Graham JA, Chen LC, Postlethwait EM, Ghio AJ, Foster WM, et al. Panel discussion review: session four--assessing biological plausibility of epidemiological findings in air pollution research. J Expo Sci Environ Epidemiol. 2007 Dec;17 Suppl 2:S97-105.

54. Esposito K, Petrizzo M, Maiorino MI, Bellastella G, Giugliano D. Particulate matter pollutants and risk of type 2 diabetes: a time for concern? Endocrine. 2016 Jan;51(1):32–7.

55. Brook RD, Newby DE, Rajagopalan S. Air Pollution and Cardiometabolic Disease: An Update and Call for Clinical Trials. Am J Hypertens. 2017 Dec 8;31(1):1–10.

56. Rao X, Patel P, Puett R, Rajagopalan S. Air pollution as a risk factor for type 2 diabetes. Toxicol Sci. 2015 Feb;143(2):231–41.

57. Liu C, Ying Z, Harkema J, Sun Q, Rajagopalan S. Epidemiological and experimental links between air pollution and type 2 diabetes. Toxicol Pathol. 2013 Feb;41(2):361–73.

58. Rao X, Montresor-Lopez J, Puett R, Rajagopalan S, Brook RD. Ambient air pollution: an emerging risk factor for diabetes mellitus. Curr Diab Rep. 2015 Jun;15(6):603.

59. Heck TG, Fiorin PBG, Frizzo MN, Ludwig MS. Fine Particulate Matter (PM2.5) Air Pollution and Type 2 Diabetes Mellitus (T2DM): When Experimental Data Explains Epidemiological Facts. In: G. AR, editor. Diabetes and Its Complications. InTech; 2018.

60. Li Y, Xu L, Shan Z, Teng W, Han C. Association between air pollution and type 2 diabetes: an updated review of the literature. Ther Adv Endocrinol Metab. 2019;10:2042018819897046.

61. Chandel S, Tiwari R, Dixit M. Endothelial Dysfunction in Diabetes. In: Kartha C, Ramachandran S, Pillai R, editors. Mechanisms of Vascular Defects in Diabetes Mellitus. Springer; 2017.

62. Schraufnagel DE, Balmes JR, Cowl CT, De Matteis S, Jung S-H, Mortimer K, et al. Air Pollution and Noncommunicable Diseases: A Review by the Forum of International Respiratory Societies' Environmental Committee, Part 1: The Damaging Effects of Air Pollution. Chest. 2019;155(2):409–16.

63. Bowen R. Physiologic Effects of Insulin [Internet]. VIVO Pathophysiology. 2019 [cited 2020 Mar 9]. Available from: http://www.vivo.colostate.edu/hbooks/pathphys/endocrine/pancreas/insulin_phys.html

64. Back SH, Kaufman RJ. Endoplasmic reticulum stress and type 2 diabetes. Annu Rev Biochem. 2012;81:767–93.

65. Gómez-Hernández A, Beneit N, Díaz-Castroverde S, Escribano Ó. Differential Role of Adipose Tissues in Obesity and Related Metabolic and Vascular Complications. Int J Endocrinol [Internet]. 2016 [cited 2020 Mar 10];2016. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5059561/

66. He D, Wu S, Zhao H, Qiu H, Fu Y, Li X, et al. Association between particulate matter 2.5 and diabetes mellitus: A meta-analysis of cohort studies. J Diabetes Investig. 2017 Sep;8(5):687–96.

67. Balti EV, Echouffo-Tcheugui JB, Yako YY, Kengne AP. Air pollution and risk of type 2 diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract. 2014 Nov;106(2):161–72.

68. Janghorbani M, Momeni F, Mansourian M. Systematic review and metaanalysis of air pollution exposure and risk of diabetes. Eur J Epidemiol. 2014 Apr;29(4):231–42.

69. Li C, Fang D, Xu D, Wang B, Zhao S, Yan S, et al. Main air pollutants and diabetes-associated mortality: a systematic review and meta-analysis. Eur J Endocrinol. 2014 Nov;171(5):R183-190.

70. Eze IC, Hemkens LG, Bucher HC, Hoffmann B, Schindler C, Künzli N, et al. Association between ambient air pollution and diabetes mellitus in Europe and North America: systematic review and meta-analysis. Environ Health Perspect. 2015 May;123(5):381–9.

71. Park SK, Wang W. Ambient Air Pollution and Type 2 Diabetes: A Systematic Review of Epidemiologic Research. Curr Environ Health Rep. 2014 Sep 1;1(3):275–86.

72. Meo SA, Memon AN, Sheikh SA, Rouq FA, Usmani AM, Hassan A, et al. Effect of environmental air pollution on type 2 diabetes mellitus. Eur Rev Med Pharmacol Sci. 2015 Jan;19(1):123–8.

73. Liu F, Chen G, Huo W, Wang C, Liu S, Li N, et al. Associations between long-term exposure to ambient air pollution and risk of type 2 diabetes mellitus: A systematic review and meta-analysis. Environ Pollut. 2019 Sep 1;252:1235–45.

74. Yang B-Y, Fan S, Thiering E, Seissler J, Nowak D, Dong G-H, et al. Ambient air pollution and diabetes: A systematic review and meta-analysis. Environ Res. 2020 Jan 1;180:108817.

75. Yang M, Cheng H, Shen C, Liu J, Zhang H, Cao J, et al. Effects of long-term exposure to air pollution on the incidence of type 2 diabetes mellitus: a meta-analysis of cohort studies. Environ Sci Pollut Res Int. 2019 Dec 6;

76. Paul LA, Burnett RT, Kwong JC, Hystad P, van Donkelaar A, Bai L, et al. The impact of air pollution on the incidence of diabetes and survival among prevalent diabetes cases. Environ Int. 2020 Jan;134:105333.

77. Requia WJ, Adams MD, Koutrakis P. Association of PM2.5 with diabetes, asthma, and high blood pressure incidence in Canada: A spatiotemporal analysis of the impacts of the energy generation and fuel sales. Sci Total Environ. 2017 Apr 15;584–585:1077–83.

78. Clark C, Sbihi H, Tamburic L, Brauer M, Frank LD, Davies HW. Association of Long-Term Exposure to Transportation Noise and Traffic-Related Air Pollution with the Incidence of Diabetes: A Prospective Cohort Study. Environ Health Perspect. 2017 31;125(8):087025.

79. To T, Zhu J, Villeneuve PJ, Simatovic J, Feldman L, Gao C, et al. Chronic disease prevalence in women and air pollution--A 30-year longitudinal cohort study. Environ Int. 2015 Jul;80:26–32.

80. Chen H, Burnett RT, Kwong JC, Villeneuve PJ, Goldberg MS, Brook RD, et al. Risk of incident diabetes in relation to long-term exposure to fine particulate matter in Ontario, Canada. Environ Health Perspect. 2013 Jul;121(7):804–10.

81. Brook RD, Cakmak S, Turner MC, Brook JR, Crouse DL, Peters PA, et al. Long-term fine particulate matter exposure and mortality from diabetes in Canada. Diabetes Care. 2013 Oct;36(10):3313–20.

82. Lim CC, Hayes RB, Ahn J, Shao Y, Silverman DT, Jones RR, et al. Association between long-term exposure to ambient air pollution and diabetes mortality in the US. Environ Res. 2018;165:330–6.

83. Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. The 2016 global and national burden of diabetes mellitus attributable to PM2·5 air pollution. Lancet Planet Health. 2018;2(7):e301–12.

84. Coogan PF, White LF, Yu J, Burnett RT, Seto E, Brook RD, et al. PM2.5 and Diabetes and Hypertension Incidence in the Black Women's Health Study. Epidemiology. 2016 Mar;27(2):202–10.

85. Park SK, Adar SD, O'Neill MS, Auchincloss AH, Szpiro A, Bertoni AG, et al. Long-term exposure to air pollution and type 2 diabetes mellitus in a multiethnic cohort. Am J Epidemiol. 2015 Mar 1;181(5):327–36.

86. Pope CA, Turner MC, Burnett RT, Jerrett M, Gapstur SM, Diver WR, et al. Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality. Circ Res. 2015 Jan 2;116(1):108–15.

87. Coogan PF, White LF, Jerrett M, Brook RD, Su JG, Seto E, et al. Air pollution and incidence of hypertension and diabetes mellitus in black women living in Los Angeles. Circulation. 2012 Feb 14;125(6):767–72.

88. Puett RC, Hart JE, Schwartz J, Hu FB, Liese AD, Laden F. Are particulate matter exposures associated with risk of type 2 diabetes? Environ Health Perspect. 2011 Mar;119(3):384–9.

89. Hansen AB, Ravnskjær L, Loft S, Andersen KK, Bräuner EV, Baastrup R, et al. Long-term exposure to fine particulate matter and incidence of diabetes in the Danish Nurse Cohort. Environ Int. 2016 May;91:243–50.

90. Renzi M, Cerza F, Gariazzo C, Agabiti N, Cascini S, Di Domenicantonio R, et al. Air pollution and occurrence of type 2 diabetes in a large cohort study. Environ Int. 2018;112:68–76.

91. Lucht S, Hennig F, Moebus S, Ohlwein S, Herder C, Kowall B, et al. All-source and source-specific air pollution and 10-year diabetes Incidence: Total effect and mediation analyses in the Heinz Nixdorf recall study. Environ Int. 2020 Mar;136:105493.

92. Weinmayr G, Hennig F, Fuks K, Nonnemacher M, Jakobs H, Möhlenkamp S, et al. Long-term exposure to fine particulate matter and incidence of type 2 diabetes mellitus in a cohort study: effects of total and traffic-specific air pollution. Environ Health. 2015 Jun 19;14:53.

93. Liang F, Yang X, Liu F, Li J, Xiao Q, Chen J, et al. Long-term exposure to ambient fine particulate matter and incidence of diabetes in China: A cohort study. Environ Int. 2019;126:568–75.

94. Lao XQ, Guo C, Chang L-Y, Bo Y, Zhang Z, Chuang YC, et al. Long-term exposure to ambient fine particulate matter (PM2.5) and incident type 2 diabetes: a longitudinal cohort study. Diabetologia. 2019;62(5):759–69.

95. Li C-Y, Wu C-D, Pan W-C, Chen Y-C, Su H-J. Association Between Long-term Exposure to PM2.5 and Incidence of Type 2 Diabetes in Taiwan: A National Retrospective Cohort Study. Epidemiology. 2019 Jul;30 Suppl 1:S67–75.

96. Qiu H, Schooling CM, Sun S, Tsang H, Yang Y, Lee RS-Y, et al. Long-term exposure to fine particulate matter air pollution and type 2 diabetes mellitus in elderly: A cohort study in Hong Kong. Environ Int. 2018;113:350–6.

97. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet. 2004 Jan 10;363(9403):157–63.

98. Yang B-Y, Qian ZM, Li S, Chen G, Bloom MS, Elliott M, et al. Ambient air pollution in relation to diabetes and glucose-homoeostasis markers in China: a cross-sectional study with findings from the 33 Communities Chinese Health Study. Lancet Planet Health. 2018;2(2):e64–73.

99. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012 Apr 1;41(2):514–20.

100. Ahlbom A, Alfredsson L. Interaction: A word with two meanings creates confusion. Eur J Epidemiol. 2005;20(7):563–4.

101. Cassee FR, Héroux M-E, Gerlofs-Nijland ME, Kelly FJ. Particulate matter beyond mass: recent health evidence on the role of fractions, chemical constituents and sources of emission. Inhal Toxicol. 2013 Dec;25(14):802–12.

APPENDICES

Appendix 1. Literature Review

Table A1.1. Cohort studies on long-term exposure to ambient $PM_{2.5}$ and developing T2DM (incidence or mortality), according to region

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall ($\mu g/m^3$)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
NORTH AMERICA									
Paul et al. (2020)	Ontario, Canada; ONPHEC; Aged 35–85 years old;	Total: 4,774,984 Cases during follow-up: 790,461	Study period: 2001–2015	Satellite-based $PM_{2.5}$ estimates derived from AOD, collected 1998–2012; Combined with CTM, and further refined with GWR; 1×1 km spatial resolution; R^2 with ground-based measurements = 0.82	Assigned as 3-year moving average; Linked using annual residential postal codes; Accounted for residential mobility	Mean: 9.6 IQR: 3.5	Incidence; Ontario Diabetes Database; ICD-9 or ICD-10 codes;	Adjusted HR for per IQR: 1.010 (1.005,1.014) Adjusted HR per 10 $\mu g/m^3$: 1.029 (1.016,1.042)	Adjusted for age, sex, SES, region, and healthcare access covariables. All models stratified by Greater Toronto Area residence Assessed potential effect modification by age group, sex, and income quintile
Requia et al. (2017)	Canada; Population-based cohort; Health regions; Aged ≥44 years	Total: 25,704,288 Cases during follow-up: 5,570,326	Study period: 2007–2014	Satellite-based $PM_{2.5}$ estimates, collected 2005–2014; Derived from AOD retrievals with CTM and calibrated using GWR; 1×1 km spatial resolution	Assigned 5-year time-weighted average $PM_{2.5}$ concentration; Linked by health region (not individual-level)	Not reported	Incidence; Canadian Community Health Survey (CCHS); self-reported	Adjusted IRR per 10 $\mu g/m^3$: 1.05 (1.02,1.13)	Adjusted for age, cohort, gender
Clark et al. (2017)	Metropolitan Vancouver, British Columbia, Canada; Population-based cohort; Aged 45–85	Total: 380,738 Cases during follow-up: 12,941	Study period: 1994–2002 Follow-up: 4 years	High-spatial-resolution LUR, collected in 2003; 10×10 km spatial resolution	Assigned 5-year time-weighted average $PM_{2.5}$ concentration; Linked using residential postal code; Accounted for annual residential mobility	Mean: 4.1 Range: 0.0–10.2 IQR: 1.6	Incidence; BC Medical Services Plan (MSP); includes Central Registry, Physician Visit, and Hospital data; ICD codes for one hospital or two physician health claims ICD-9 or ICD-10 (ICD-9 code 250, and ICD-10	Adjusted OR per IQR: 1.03 (1.01,1.05) Adjusted OR per 10 $\mu g/m^3$: 1.20 (1.07,1.36)	Adjusted for age, gender, and area-level household income

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall ($\mu g/m^3$)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
	years old						coding back translated to ICD-9 coding)		
To et al. (2015)	Ontario, Canada; Canadian National Breast Screening Study (CNBSS); RCT for screening breast cancer; Longitudinal cohort; Aged 40–59 years old; Women	Total: 29,549 Cases during follow-up: Unclear	Study period: 1992 1996–2013	Satellite-based $PM_{2.5}$ estimates derived from AOD from MODIS & MISR instruments, collected 1998–2006; 10×10 km spatial resolution; Correlation with fixed-site ground measurements (r=0.77, n=1057)	Average, time-fixed; Linked using residential postal code at baseline only; Had data on whether participants ever moved from baseline residence	Not reported	Incidence; Baseline questionnaires; administrative data; including OHIP (physician outpatient claims); NACRS (emergency visits); and CIHI DAD (hospital discharges); ICD-9 or ICD-10 codes; Prevalent cases after 1996 considered as incidence;	Adjusted IRR per 10 $\mu g/m^3$: 1.28 (1.13,1.45)	Adjusted for age, education, occupation, marital status, smoking, BMI and contextual variables derived from census area measures
Chen et al. (2013)	Ontario, Canada; Population-based cohort; Based on 5 cycles of surveys; Aged ≥35 years old	Total: 62,012 Cases during follow-up: 6,310	Study period: 1996–2010 Mean ±SD follow-up: 8 ±3.2 years	Satellite-based $PM_{2.5}$ estimates derived from AOD from MODIS & MISR instruments, collected 2001–2006; 10×10 km spatial resolution; Correlation with fixed-site ground measurements (r=0.77, n=1057)	Assigned 6-year average $PM_{2.5}$ concentration; Linked using residential postal code at baseline only	Mean: 10.6 Range: 2.6–19.1	Incidence; Ontario Diabetes Database; includes CIHI DAD; OHIP ICD-9 code 250; ICD-10 codes E10–E14; ≥1 hospital admission with diabetes diagnosis; or ≥2 physician claims for diabetes within 2-year period; Validated registry and algorithm;	IR per 1,000 person-years: 13.0 Adjusted HR per 10 $\mu g/m^3$: 1.11 (1.02,1.21)	Adjusted for sex, marital status, education, household income adequacy, BMI, physical activity, smoking, alcohol consumption, diet, race, hypertension, urban residency, area-level unemployment rate, education, and household income; Stratified by age, cycle of survey, and region (south/north)
Brook et al. (2013)	Canada; Population-based	Total: 2,145,400 Cases	Study period: 1991–	Satellite-based $PM_{2.5}$ estimates derived from AOD, collected 2001–	Assigned 6-year average $PM_{2.5}$ concentration;	Mean ±SD: 8.7 ±3.9 IQR: 6.2	Mortality; Canadian Mortality Database;	Adjusted HR per 10 $\mu g/m^3$: 1.49 (1.37,1.62)	Adjusted for Aboriginal ancestry, visible minority,

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
	cohort; Canadian Census Mortality Follow-up Study; Aged ≥25 years old	during follow-up: 5,200	2001 Follow-up: <10 years	2006, historical estimates for 1987–2001; 10 × 10 km spatial resolution; Correlation (r=0.89)	Linked using residential postal code at baseline only		ICD-9 code 250 or ICD-10 codes E10–E14		marital status, education, employment status, occupation, low income cut-off, community size, and area-level contextual covariates; Stratified by age, sex
Lim et al. (2018)	USA; NIH-AARP Diet and Health Study; Aged 50–71 years old	Total: 549,735 Cases during follow-up: 3,598	Study period: 1995–2011	EPA monitoring	Assigned average (2002–2010), time-fixed; Sensitivity analysis by assigning time-varying 1-year lagged exposure	Mean ±SD: 11.0 ±2.7 Range: 2.8–21.2	Mortality; Social Security Administration Death Master File for T2DM mortality; Mailed questionnaire for covariates	Adjusted HR per 10 μg/m³: Main analysis 1.19 (1.03,1.39) Time-varying 1.18 (1.03, 1.36)	Adjusted for age, sex, and location as strata; and race, BMI, education, smoking, and marriage at individual-level; and median income and % with high school education at census-tract level Assessed effect modification by age, sex, race, smoking history, pre-diagnosed health status, BMI and diet Evidence of effect modification by BMI (higher) and fruit consumption (lower)
Bowe et al. (2018)	USA; Longitudinal cohort; US veterans; ~94% men	Total: EPA cohort 1,729,108 Satellite cohort 1,670,031 Cases during follow-up: 397,966	Study period: 2003–2012 Median follow-up: 8.5 years	EPA AQS monitors within 30, 10, & 5 miles; Satellite-based $PM_{2.5}$ estimates derived from AOD from MODIS, MISR, & SeaWIFS instruments as alternate method	Assigned at baseline, and as time-varying variable; Linked using residential postal code; Accounted for annual residential mobility	Mean: 11.8 Range: 5.0–22.1 IQR: 7	Incidence; US Department of Veterans Affairs' (VA) databases; includes routine blood panels from VA Healthcare System; ICD-9 code, diabetes medication prescription, or HbA1c >6.4% (>46.4	IR per 1,000 person-years: 34.1 Adjusted HR per 10 μg/m³: Baseline EPA 1.15 (1.08,1.22) Time-varying EPA	Adjusted for age, race, sex, estimated GFR, systolic blood pressure, hyperlipidemia, chronic lung disease, cardiovascular disease, cancer, BMI, smoking status, angiotensin-

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
							mmol/mol)	1.18 (1.10,1.25) Satellite 1.13 (1.11,1.15)	medicationg, percentage of people in poverty in each county of residence, population density of county of residence, number of admissions to hospital before beginning of follow-up, and how many times serum creatinine was measured before beginning of follow-up
Park et al. (2015)	USA; Multi-Ethnic Study of Atherosclerosis (MESA); Aged 45–84 years old; Black, white, Chinese, & Hispanic	Total: 5,135 Cases during follow-up: 622	Study period: 2002–2012 Follow-up: 9 years	Hierarchical spatiotemporal model using collections from EPA AQS, supplemental home and within community measurements, and LUR covariates; Collected for year 2000	Assigned as annual average at baseline; Linked using residential postal code at baseline only; But addressed residential mobility	Mean: 16.7 IQR: 2.43	Incidence; Clinical examination; sample measure; questionnaire; calibrated devices; Doctor diagnoses; taking diabetes medication; or FPG ≥126 mg/dL	IR per 1,000 person-years: 15.9 Adjusted HR per IQR: 1.02 (0.95,1.11) Adjusted HR per 10 μg/m³: 1.08 (0.79,1.49)	Adjusted for age, sex, race, family history of diabetes, educational, smoking status, alcohol consumption, physical activity, neighborhood SES, and BMI; Also evaluated effect modification by sex, race/ethnicity, educational level, and obesity
Pope et al. (2015)	USA; American Cancer Society Cancer Prevention Study II Cohort (ACS CPS-II); Aged ≥30 years	Total: 669,046 Cases during follow-up: 4,890	Study period: 1982–2004 Follow-up: 22 years	Hybrid LUR and BME model; Monitoring from 1,464 sites, collected 1999–2008; Step 1 LUR, step 2 BME kriging interpolation; Training data set of 1,329 monitors for cross-validation (R^2 = 0.79)	Assigned as average of estimates from 1999–2004, time-invariant; Linked using residential address at baseline only;	Mean ±SD: 12.6 ±2.9 Range: 1–28	Mortality; National Death Index (database); and baseline questionnaire; Biennial vital status from death certificate using ICD-9 250 and ICD-10 E10–E14	Adjusted HR per 10 μg/m³: 1.13 (1.02,1.26)	Adjusted for current and former smoking habits; second-hand cigarette smoke; workplace PM2.5 exposure in each subject's main lifetime occupation; self-reported exposure to dust and fumes in the workplace; marital status; level of

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
									education; BMI; consumption of alcohol; dietary fat index and dietary vegetable/fruit/fiber index, median household income; percentage of people with <125% of poverty-level income; percentage of unemployed individual aged ≥16 years; percentage of adults with <12th grade education; and percentage of the population who were black or Hispanic Stratified by age, sex, race
Puett et al. (2011) (NHS cohort)	USA; Nurses' Health Study (NHS); Women, aged 30–55;	Total: 74,412 Cases during follow-up: 3,784	Study period: 1976 1989–2002 Follow-up: 13 years	GIS-based spatiotemporal models, using monitor data from EPA AQS, VIEWS, IMPROVE network, Stacked Filter Unit, CASTNet networks, and LUR covariates; Monitoring data from 498 sites	Assigned as annual time-varying exposure, 1-year lag; Linked using residential postal code at baseline only; No data on residential mobility	Mean ±SD: 17.5 ±2.7 IQR: 4.30	Incidence; Biennial mailed questionnaires; thus self-reported; National Diabetes Data Group criteria; such as elevated plasma glucose concentrations (FPG, random plasma glucose concentration, 2-h OGTT); symptoms; or self-report hypoglycemic medication; Validated with medical records;	IR per 1,000 person-years: 4.48 Adjusted HR per 10 μg/m³: 1.05 (0.87,1.22)	Adjusted for age, season, calendar year, state of residence, time-varying cigarette smoking, time-varying hypertension, baseline BMI, time-varying alcohol intake, baseline physical activity, time-varying diet
Puett et al. (2011)	USA; Health	Total: 15,048	Study period:	GIS-based spatiotemporal models,	Assigned as annual time-varying	Mean ±SD: 18.3 ±3.1	Incidence; Biennial mailed	IR per 1,000 person-years:	Adjusted for age, season, calendar year,

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
(HPFS cohort)	Professionals Follow-up Study (HPFS); Men, aged 40–75	Cases during follow-up: 688	1986 1989–2002 Follow-up: 13 years	using monitor data from EPA AQS, VIEWS, IMPROVE network, Stacked Filter Unit, CASTNet networks, and LUR covariates; Monitoring data from 498 sites	exposure, 1-year lag; Linked using residential and sometimes work postal codes at baseline only; No data on residential mobility	IQR: 4.00	questionnaires; thus self-reported; National Diabetes Data Group criteria; such as elevated plasma glucose concentrations (FPG, random plasma glucose concentration, 2-h OGTT); symptoms; or self-report hypoglycemic medication; Validated with medical records;	4.02 Adjusted HR per 10 μg/m³: 1.18 (0.81,1.71)	state of residence, time-varying cigarette smoking, time-varying hypertension, baseline BMI, time-varying alcohol intake, baseline physical activity, time-varying diet
Coogan et al. (2012)	Los Angeles, USA; Black Women's Health Study (BWHS); African American women; Subscribers to *Essence* magazine; Aged 21–69 years old	Total: 3,992 Cases during follow-up: 183	Study period: 1995–2005 Mean follow-up: 10 years	Interpolation with kriging model from 23 state and local district monitoring station in 2000	Assigned as annual time-varying based on measures in 2000; Linked using residential postal code; Had data on whether participants ever moved from baseline residence	Mean ±SD: 20.7 ±2.1	Incidence; Biennial mailed questionnaire; Self-report doctor-diagnosis at age >30 years (to exclude cases of type 1); Validation study with confirmed with medical records provided by physicians	IR per 1,000 person-years: 5.4 Adjusted IRR per 10 μg/m³: 1.63 (0.78,3.44)	Adjusted for age, BMI, income, number of people in the household, history of diabetes, smoking, physical activity, and neighborhood SES score
Coogan et al. (2016)	56 metro areas, USA; BWHS; African American women; Aged 21–69 years old	Total: 43,004 Cases during follow-up: 4,387	Study period: 1995–2011	Hybrid modeling that incorporated a spatiotemporal LUR and BME approach; EPA AQS of 1,464 monitoring locations; Collected 1999–2008	Assigned as average of estimates from 1999–2008, time-invariant; Linked using residential postal code; Accounted for residential mobility	Mean: 13.9 IQR: 2.9	Incidence; Biennial mailed questionnaire; and web-based health questionnaires; Self-report doctor-diagnosis at age >30 years (to exclude cases of type 1); Validation study with	IR per 1,000 person-years: 9.7 Adjusted HR per IQR: 1.13 (1.04,1.24) Adjusted HR per 10 μg/m³: 1.52 (1.13,2.06)	Adjusted for neighborhood SES, BMI, education, vigorous exercise, diet pattern Stratified by age, questionnaire cycle, metro area

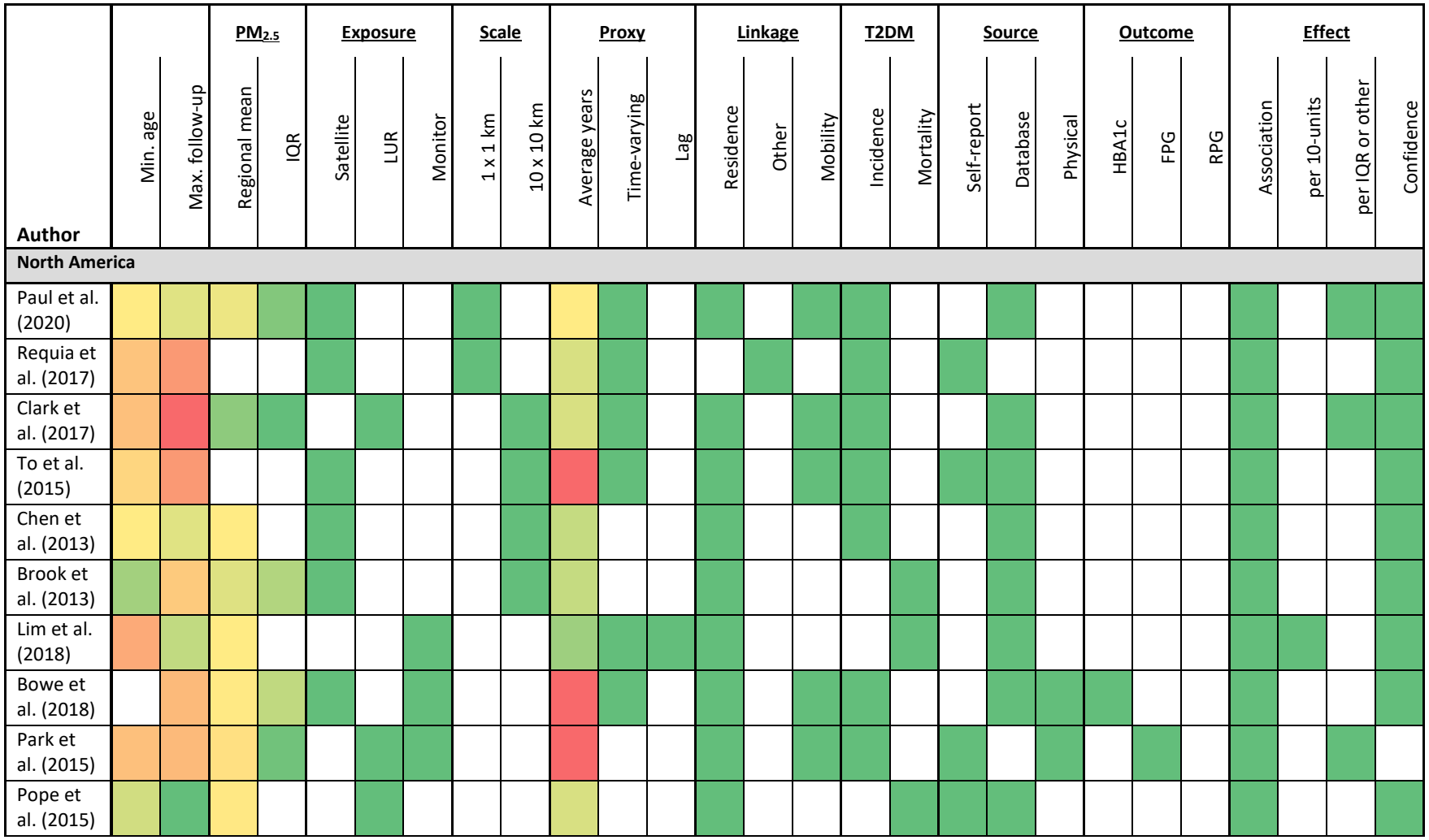
Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
							confirmed with medical records provided by physicians		
EUROPE									
Hansen et al. (2016)	Denmark; Danish Nurse Cohort (DNC); Female, aged ≥44 years Analytical cohort: Mean age: 54.0 years	Total: 24,174 Cases during follow-up: 1,137	Study period: 1993–2012 Mean follow-up: 15.3 years	GIS-based dispersion model; High-resolution Danish air pollution dispersion modeling system (AirGIS)	Assigned as 5-year moving average exposure; Linked at residential address; Accounted for annual residential mobility	Mean ±SD: 18.1 ±2.8 IQR: 3.1	Incidence; National Diabetes Register; Diabetes hospital discharge diagnosis for ICD10 code E10–14, DH36.0, DO24; chiropody as a diabetic patient; blood glucose measures; or diabetes medication purchase	IR per 1,000 person-years: 3 Adjusted HR per IQR: 1.11 (1.01,1.22) Adjusted HR per 10 μg/m³: 1.40 (1.03,1.90) C-R shape: linear	Age, calendar time, smoking, physical activity, alcohol consumption, fatty meat consumption, fruit and vegetable consumption, employment status, marital status, BMI, hypertension and MI; Effect modification assessment by age, physical activity, BMI, smoking status, MI, hypertension, and level of urbanization Positive effect modifiers: non-smoker, obese, CVD
Renzi et al. (2018)	Italy; Rome Longitudinal Study (RoLS); Aged ≥35 years Analytical cohort: Female: 55%	Total: 1,319,193 Cases during follow-up: 65,955	Study period: 2008–2013 Mean follow-up: 5.2 years	LUR, developed using measures in 2010 for ESCAPE project; 20 sampling sites	Linked using residential address at baseline only	Mean: 19.6 IQR: 1.7	Incidence; Registry; Qualified for free diabetes healthcare (code: 013.250); diagnosis (code: 250.0); prescribed hypoglycemic drugs (ATC code: A10); Validated with HbA1c ≥5.7% random population samples	IR per 1,000 person-years: 9.6 IR per 1,000 person-years, standardized for age and sex: 10.1 Adjusted HR per 5-μg/m³: 1.00 (0.97,1.02) Adjusted HR per 10 μg/m³: 0.99 (0.95,1.04)	Adjusted for SES, marital status, educational level, occupation, place of birth and sex, LDEN, NDVI, pre-existing comorbidities Assessed potential effect modification by age, sex, education, and previous comorbidities
Weinmayr	Ruhr area,	Total:	Study	European Air Pollution	Assigned as 2-year	Mean:	Incidence;	Adjusted RR per	Adjusted for age,

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
et al. (2015)	Germany; Heinz Nixdorf Recall Study (HNRS); Aged 45–75 years	3,607 Cases during follow-up: 331	period: 2000–2008 Mean follow-up: 5.1 years	Dispersion and Chemistry Transport Model (EURAD- CTM); 1 × 1 km spatial resolution	average (2001–2002) exposure, time-invariant; No 2003 due to great heat wave in Europe; Linked using residential address at baseline only	16.7 IQR: 2.29	Self-administered questionnaire; face-to-face interviews; clinical examination including lab tests; Self-reported physician diagnosis; taking diabetes medication with ATC code A10; RPG ≥200 mg/dL; or FPG ≥126 mg/dL	1-μg/m³: 1.03 (0.95,1.12) Adjusted RR per 10 μg/m³: 1.34 (0.59,3.06)	gender, lifestyle variables, BMI, individual and neighborhood SES, and city; Effect modification was investigated for age, sex, physical activity, BMI. Education, smoking status, and hs-CRP Positive effect modifiers: male, obese
Lucht et al. (2020)	Ruhr area, Germany; Heinz Nixdorf Recall Study (HNRS); Aged 45–75 years	Total: 2,451 Cases during follow-up: 237	Study period: 2000/2003–2011/2015 Mean follow-up: 10.3 years	European Air Pollution Dispersion and Chemistry Transport Model (EURAD- CTM); 1 × 1 km spatial resolution	Assigned as 2-year average (2001–2002) exposure, time-invariant	Median: 17.1 IQR: 2.2	Incidence; Self-reported physician diagnosis; or taking diabetes medication	Adjusted RR per 1-μg/m³: 1.06 (0.98, 1.16) Adjusted RR per 10 μg/m³: 1.34 (0.58, 3.12)	Adjusted for age, chronic traffic noise exposure, sex, alcohol consumption, physical activity, smoking status, ETS exposure, cumulative smoking, and nutrition. Effect modification was investigated for sex.
ASIA									
Liang et al. (2019)	China; China-PAR Project, includes 4 large-scale cohorts; Aged 35–74 years Analytical cohort: Mean age: 51.7 ±11.7 years	Total: 88,397 Cases during follow-up: 6,439	Study period: 2004–2015; Follow-up: 11 years	Satellite-based $PM_{2.5}$ estimates derived from AOD from MODIS instrument, collected 2004–2015; Deep Blue & Dart Target algorithm used to improve data coverage; Meteorological data extracted from ECMWF ERA-Interim; 10 × 10 km spatial resolution;	Assigned as time-weighted average over full follow-period; Linked using residential postal code;	Mean ±SD: 79.1 ±13.8	DM Incidence; Standardized questionnaire interview; physical examination; and blood sampling; FPG ≥7.0 mmol/L, use of insulin or hypoglycemic agents; or diagnosed medical history of diabetes	IR per 1,000 person-years: 11.1 Adjusted HR per 10 μg/m³: 1.16 (1.06,1.26)	Age, gender, BMI, smoking, education, work-related physical activity, hypertension, urbanicity, education, and temperature and relative humidity Stratified analyses for age, gender, urbanicity, BMI, smoking status, and hypertension; Additional effect

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
	Female: 60.2%								modification assessment with gender-specific smoking Positive effect modifiers: younger, female, rural, lower BMI, non-smokers, non-hypertensive
Lao et al. (2019)	Taiwan, China; Prospective cohort; Private firm MJ Health Management Institution, paid membership, at last 2 medical visits; Aged ≥18 years Analytical cohort: Mean age: 38.3 ±11.5 years Female: 49.9%	Total: 147,908 Cases during follow-up: 4,781	Study period: 2001–2014 Mean follow-up: 6.7 years Follow-up range: 2–13.9 years	Satellite-based $PM_{2.5}$ estimates derived from AOD from MODIS instrument; Correction factor using ground observations; Calibrated using AERONET in Taipei, Taiwan; Validated with measures from 70 ground-level monitors; 1 × 1 km spatial resolution;	Assigned as 2-year moving average exposure, no lag; Linked using either residential or business address; Accounted for residential mobility per visit interval (mean ±SD was 1.2 ±4.1 years)	Mean ±SD: 26.5 ±7.4 IQR: 6.3 Q1: <21.7 Q2: 21.7–<24.1 Q3: 24.1–<28.0 Q4: ≥28.0	T2DM Incidence; Medication examination; standard self-administered questionnaire; Self-reported physician-diagnosis; or FPG ≥7 mmol/l	IR per 1,000 person-years: 3.5 Adjusted HR for Q4 vs Q1: 1.16 (1.07,1.26) Adjusted HR for Q3 vs Q1: 1.27 (1.17,1.38) Adjusted HR for Q2 vs Q1: 1.28 (1.18,1.39) C-R shape: non-linear	Adjusted for age, sex, education, season, year, smoking status, alcohol drinking, physical activity, vegetable intake, fruit intake, occupational exposure, BMI, hypertension and dyslipidaemia (all were treated as time-dependent covariates except for sex); Stratified analysis for sex, education, smoking status, alcohol drinking, physical activity, BMI, hypertension, and dyslipidaemia Positive effect modifiers: lower BMI, alcohol drinker
Li et al. (2019)	Taiwan, China; Retrospective cohort; Taiwanese National Health	Total: 505,151 Cases during follow-up: 48,611	Study period: 2001–2012 Median follow-up: 12	Monitoring data from EPA of Taiwan from 2006–2012; LUR;	Assigned as 7-year average (2006–2012); Linked using participant's most frequently visited medical facility at	Mean: 27.92 Range: 13.81–39.94 Median: 27.28	Incidence; Longitudinal Health Insurance Database (LHID) 2000; ICD-9 (code 250) for more than 3 times in 1 year; concurrent anti-	12-year cumulative incidence: 8.4% Adjusted HR per 10 μg/m³: 1.11 (1.08,1.13) C-R shape: linear	Adjusted for individual variables including age, sex, National Health Insurance premium, occupational type, hypertension, and

Author	Study: Participants	Sample Size	Period	$PM_{2.5}$ Exposure: Model or Monitor	Assigned; Linked; Mobility	Overall (μg/m³)	T2DM Outcome: DM; Measure; Source; Definition	Effect estimate (95% CI)	Adjustments
	Insurance Research Database (NHIRD); Aged ≥20 years Analytical cohort: Mean age: 42.6 ±15.8 years Female: 51.3%		years		baseline;		hyperglycemic medication		hyperlipidemia, and contextual socioeconomic circumstances (area-level covariates) including county-level income and township-level urbanization; Examined effect modification of sex, age, hyperlipidemia, and National Health Insurance premium Positive effect modifiers: male, older, hyperlipidemia
Qiu et al. (2018)	Hong Kong, China; Elderly Health Services (EHS) Cohort; Volunteers, elderly, aged ≥65 years Analytical cohort: Mean age: 72 years Female: 65.9%	Total: 53,905 Cases during follow-up: 806	Study period: 1998–2010 Mean follow-up: 9.8 years	Satellite-based $PM_{2.5}$ estimates derived from AOD, collected 1998–2010; Surface extinction coefficients (SEC); 1×1 km spatial resolution; Calibrated against surface measurements from 4 stations;	Assigned as annual time-varying exposure; Linked using residential postal code; Accounted for annual residential mobility	Mean: 35.8 IQR: 3.2	Incidence; Standardized and structured interview; and clinical examination; Hospital Authority records; Hospital discharge records using ICD-9 codes 250x0 and 250 x2 (x=0–9)	IR per 1,000 person-years: 1.5 Adjusted HR per IQR: 1.15 (1.05,1.25) Adjusted HR per 10 μg/m³: 1.55 (1.18,2.03) C-R shape: linear	Adjusted for age, sex, BMI, smoking status, alcohol drinking, physical exercise, education, monthly expenses, medication taken, and self-reported active comorbidities, TPU level covariates (prevalence of age ≥65, tertiary education and income ≥US$ 1923/m) and smoking rate at district level; Stratified by sex Positive effect modifier: female

Table A1.2. Heat map of cohort studies



Author	Min. age	Max. follow-up	$PM_{2.5}$		Exposure			Scale		Proxy			Linkage			T2DM		Source			Outcome			Effect			
			Regional mean	IQR	Satellite	LUR	Monitor	1 x 1 km	10 x 10 km	Average years	Time-varying	Lag	Residence	Other	Mobility	Incidence	Mortality	Self-report	Database	Physical	HBA1c	FPG	RPG	Association	per 10-units	per IQR or other	Confidence
Puett et al. (2011) (NHS)																											
Puett et al. (2011) (HPFS)																											
Coogan et al. (2012)																											
Coogan et al. (2016)																											
Europe																											
Hansen et al. (2016)																											
Renzi et al. (2018)																											
Weinmayr et al. (2015)																											
Lucht et al. (2020)																											
Asia																											
Liang et al. (2019)																											
Lao et al. (2019)																											

Author	Min. age	Max. follow-up	$PM_{2.5}$		Exposure			Scale		Proxy			Linkage			T2DM		Source			Outcome			Effect			
			Regional mean	IQR	Satellite	LUR	Monitor	1 x 1 km	10 x 10 km	Average years	Time-varying	Lag	Residence	Other	Mobility	Incidence	Mortality	Self-report	Database	Physical	HBA1c	FPG	RPG	Association	per 10-units	per IQR or other	Confidence
Li et al. (2019)																											
Qiu et al. (2018)																											

Appendix 2. Rural Deqing Cohort

Table A2.1. Timeline of Covariate Assessment in Rural Deqing Cohort

Covariates	Baseline	Follow-up (Year) 3	5	6	7	8	9
Questionnaire:							
General information	✓						
Medical history	✓						
Lifestyle & behavioural factors	✓						
Health services	✓						
Family history	✓						
Medical examination:							
Anthropomorphic measurements	✓						
Blood tests (including FPG)	✓	✓	✓	✓	✓	✓	✓
Self-reported (physician diagnosis of T2DM or anti-diabetic medication)	✓	✓	✓	✓	✓	✓	✓

Abbreviations: FPG, fasting plasma glucose; T2DM, type 2 diabetes mellitus.

Paper A2.1. Informed Consent (English Translation)

Research title: Cohort Study on Type 2 Diabetes of Adults in Deqing Rural Community
Project Leader: Chaowei Fu, School of Public Health, Fudan University
Tel: 021-54237811

You are invited to participate in this project. This informed consent provides relevant information for this study. If you agree to participate, you will be asked to sign the informed consent.

1. Research purpose
The Fudan University School of Public Health and the University of Ottawa launched the study. The study sought to understand the epidemiological status of adult diabetes and the incidence of different levels of glucose tolerance in rural communities in China, and to identify the non-genetic risk factors for adult type 2 diabetes in rural communities in China; in combination with nested case-control studies and pedigree studies, looking for possible new Susceptible genes (haplotypes), describing possible genetic patterns of type 2 diabetes, provide insight into the possible genetic etiology of type 2 diabetes and the role of gene-environment and gene-gene interactions in the development of type 2 diabetes.
The study was funded by the National Natural Science Commission. We hope that the information in this study will be helpful for the subsequent development of effective diabetes prevention programs or the development of related policies.

2. Research process
If you decide to participate in this study, you will need to take a 15-minute survey first. You will be asked about your personal background as well as your current health condition, lifestyle, etc., and take 5ml of venous blood for blood glucose testing and related genetic testing. After that, we will also conduct an identical investigation on you for a period of about 2 years.

3. Possible risks
You may worry that your privacy or personal information will leak. We will ask everyone who

participated in the study to sign a confidentiality agreement to ensure that our researchers will not disclose your relevant information.

4. Possible benefits
You will receive our free medical check-ups regularly and will always provide you with information and advice on health and diabetes prevention.

5. Subsidy to participate in this study
Every time you participate in the survey, you will receive a living allowance to compensate for your time and express our gratitude.

6. Confidentiality of research records
All research information will be locked in the file cabinet and it is absolutely confidential. Only project researchers can access this information.

7. Right to study
You can withdraw from this study at any time without any consequences. Participating in this study will not harm any of your existing rights. If you have any questions about this study, you can contact Chaowei Fu, the person in charge of the project, at 021-54237811. You can also contact the Ethics Committee of the School of Public Health of Fudan University at 021-54237051, which represents your interests.

8. Signed or authorized signature after informed consent
I have read or others have read to me and have understood the above information and content. All my questions have been satisfactorily answered.

I am willing to participate in the research mentioned in this article
Subject's signature: _____ Date _____
or Signature of the subject's authorizer: _____ Date _____

9. Investigator Statement
I have explained the study to participants and answered all the questions. I believe he decided to participate in this study.

Investigator Signature: _____ Date _____

Paper A2.2. Questionnaire (English Translation)

Hello! In order to build a new socialist countryside and show the government's concern for the health of the people, we will carry out the “Deqing Health” project and establish the health records of residents in Deqing County. This project needs your cooperation. Under normal circumstances, it may take about 20 minutes to complete your health investigation. All information in the questionnaire will be kept strictly confidential and used only for the evaluation of health. Your personal information, such as name and address, will be kept at the Deqing County Center for Disease Control and Prevention. If you have any questions about this survey, please contact the investigator. If you have further questions, please contact the Deqing County Center for Disease Control and Prevention (Tel: 8820885). If you understand the purpose of our investigation, please sign below:

I voluntarily participated in the “Deqing Health” project of the Deqing County Center for Disease Control and Prevention. All my questions were answered and I understood the significance of the survey information. This information will only be used to evaluate health.

Signature: _____ Date: _____year ____month ____day

Resident Health Questionnaire

Study ID: _____ (Do not fill in)
Family ID: _____ Personal ID:_____
Name: _____
Family Address: _____ Township (town) _____ Village
_____ Natural Village _____ Group
Telephone: _____ ID: _____

1. General situation

1.1 Gender:
(1) Male (2) Female

1.2 Birthday: _____ year ____month ___day (Lunar calendar, please specify)

1.3 Nationalities:
(1) Han (2) Others: _____

1.4 Education:
(1) Illiteracy (2) Not-graduated from elementary school
(3) Graduation from elementary school (4) Graduation from junior high school
(5) Graduation from high school (including technical secondary school, vocational high school)
(6) College degree and above

1.5 Major occupations:
(1) Farming (2) Public Officers (3) Corporate Technology/Management
(4) Workers/Private Employees (5) Business (6) Others: _____

1.6 Current working status:
(1) On-job (2) Unemployed/Laid-off (3) Retired (4) Illness

(5) Others: _____

1.7 Household size: _____ People

1.8 The average annual income of the person: _____yuan/year (Note: 5 yuan = 1 canadian dollar)

1.9 Average household annual income: _____yuan/year

1.10 Your family's local economic level:
(1) Lower (2) Medium (3) Higher (4) Wealthy

1.11 Your medical insurance is (multiple choices)
(1) Social health insurance (public expense, labor insurance, etc.)
(2) Rural cooperative medical care (3) Commercial medical insurance
(4) None (5) Don't know

1.12 You accumulatively lived in your home country for exactly _____ years. In the past 5 years, you have been living there on average for _____ months/years.

1.13 The type of house your family currently lives in is a:
(1) Self-built brick house (2) Western style sing house (3) Apartment
(4) Muddy house (5) Thatched house (6) Other: _____

1.14 The house your family currently lives in is:
(1) Owned (2) Private leasing (3) Public housing (4) Free use

1.15 There are _____ square meters of property owned by your family. There are a total of _____ square meters of houses that you live in, and you stay in _____ years.

1.16 Your current housing ventilation:
(1) Good (2) Normal (3) Poor

1.17 Your current housing lighting:
(1) Good (2) Normal (3) Poor

1.18 What relatively expensive home appliances do you have in your home:
(1) Color TV (2) Refrigerator (3) Air conditioner (4) Electric fan
(5) Recorder (6) Washing Machine (7) Water Heater (8) DVD Player
(9) Others _____

1.19 Your marital status:
(1) Never married (go to question 2.1)
(2) Married (non-remarriages) (3) Remarriage (4) Divorced (5) Widowed

1.20 First marriage time: _____ years

1.21: Times of pregnancies: _____ times, times of live birth: _____ times (please, only women answer)

1.22 Existing children: number of males _____, number of females _____

1.23 Contraceptive measures (multiple choices):
(1) Ligation (2) Relief (3) Medication (4) Condoms (5) None

2. Disease related

2.1 Have you ever had viral hepatitis?
(1) Yes (2) No (go to question 2.2) (3) Don't know (go to question 2.2)

2.1.1 Have you ever had liver disease?
(1) Yes, confirmed time _____ years (2) No (3) Don't know

2.1.2 Have you ever had Hepatitis B?
(1) Yes, confirmed time _____ years (2) No (3) Don't know

2.2 Have you had other hepatitis?
(1) Yes, specific: _____, confirmed time _____ years (2) No (3) Don't know

2.3 Has a doctor ever diagnosed you with cirrhosis?
(1) Yes, confirmed time _____ years (2) No (3) Don't know

2.4 Has a doctor ever diagnosed you with diabetes?
(1) Yes, confirmed time _____ years, is _____ type (2) No (3) Do not know

2.4.1 Have you ever measured blood sugar levels in a hospital?
(1) Yes (2) No (go to question 2.4.3) (3) Don't know (go to question 2.4.3)

2.4.2 If yes, the doctor said that your blood glucose is:
(1) high (2) normal (3) low (4) do not know

2.4.3 In the most recent year, did you have the following symptoms? (Multiple choice)
(1) Thinner than before (2) Eat more than before
(3) Drink more than before (4) Urinate more than before
(5) Fatigue more easily than before (6) More easily infected than before
(7) Less likely to heal than previous wounds

2.5 Did you ever have a doctor diagnose you with any of the following illnesses?
(1) Tuberculosis (2) Chronic bronchitis or emphysema (3) Asthma
(4) Stroke (5) Gallstones (6) Gastritis
(7) Hypertension (8) Coronary heart disease (9) Tumor, specific _____
(10) Other diseases, specifically _____ (11) No more diseases (12) Don't know

2.6 Do you usually snore?
(1) Yes (2) No (3) Don't know

3. Behavior-related factors

3.1Have you ever smoked at least one cigarette per day for more than 6 months?
(1) Yes (2) No (go to question 3.6)

3.2 You started smoking when you were _____ years old?

3.3 So far, you have smoked for _____ accumulated years?

3.4 Have you completely quit smoking as of now?
(1) Yes, quit at age _____ years, _____ months (go to question 3.6) (2) No

3.5 How often do you smoke and how much?
(1) Every day, an average of _____pieces/day
(2) Occasionally, usually _____ days/week, an average of _____pieces/day

3.6 Now, do you often have people around you who smoke?
(1) Yes (2) No (go to question 3.12)

3.7 Do you estimate that there are _____ months in the last year?

3.8 Is there someone at home who smokes around you?
(1) Yes (2) No (go to question 3.10)

3.9 You estimate that your family smokes around you
(1) Daily, approximately _____pieces/day, approximately _____ minutes/day
(2) Occasionally, usually _____days/week, _____pieces/day, approximately _____minutes/day

3.10 Do friends often smoke around you?
(1) Yes (2) No (go to question 3.12)

3.11 You estimate that a friend is smoking by your side
(1) Daily, approximately _____pieces/day, approximately _____ minutes/day
(2) Occasionally, usually _____days/week, _____pieces/day, approximately _____minutes/day

3.12 Do you often drink beer, fruit wine (wine), liquor, rice wine or any other alcoholic beverage?
(1) Yes (2) No (Go to question 3.18)

3.13 What you often drink is: (can choose more than one)
(1) Beer (2) Rice wine (3) Liquor (4) Wine (5) Other alcoholic drinks _____

3.13.1 You have been drinking for a total of _____ years

3.14 Did you drink beer in the last year?
(1) Yes, average ___ times/week, ___ bottle/time (2) No

3.15 Did you drink rice wine in the last year?
(1) Yes, average ___ times/week, ___two/times (2) No

3.16 Did you drink liquor in the last year?
(1) Yes, average ___ times per week, ___two times/times (2) No

3.17 Did you drink fruit wine in the last year?
(1) Yes, average ___ times/week, ___two/times (2) No

3.18 What kind of water do you mainly drink now?
(1) Tap water (2) Well water (3) Pond water (4) River water
(5) _____ water

3.19 Do you often come into contact with pesticides?
(1) Yes, specific name: _____ (2) No (go to question 3.20)

3.19.1 Are you often exposed to pesticides without protective measures?
(1) Yes, specific name: _____ (2) No

3.20 What fuel does your family use to cook and boil? (Multiple choice)
(1) Gas/Natural gas (2) Coal fired (3) Wood/Bamboo (4) Others: _____

3.21 Did you exercise in the last year?
(1) Yes (2) No (Go to question 3.22)

3.21.1 The frequency of your general exercise is:
(1) 7+ times/week (2) 4-6 times/week (3) 2-3 times/week
(4) 1 time/week (5)<1 time/ week

3.21.2 The duration of your exercise is generally continuous:
(1) <20 minutes (2) 20-39 minutes (3) 40-59 minutes (4) >=60 minutes

3.21.3 Your main training methods are:
(1) Running (2) Brisk walking (3) Play ball (4) Martial arts
(5) Dancing (6) Others: _____

3.22 The taste of your diet tends to:
(1) Light (2) Saltier (3) Sweeter (4) Spicy
(5) Others: _____

3.22.1 The main cooking method of your home cooking is (multiple choices)
(1) Fried (2) Cooking (3) Frying (4) Marinated
(5) Others: _____

3.22.2 General aspects of your home eating:
(1) Vegetable-based (2) Meat-based (3) Vegetables and meat

(4) Others: _____

3.22.3 What diet do you prefer to eat?
(1) Hot (2) Warm (3) Cold (4) Other: _____

3.22.4 Usually, do you eat more than average people?
(1) Fast (2) Average (3) Slow

3.22.5 Is your diet regular?
(1) Usually eat on time (2) Occasionally do not eat on time (3) Do not eat regularly on time

3.22.6 Do you prefer to eat vegetables or meat?
(1) Vegetables (2) Meat (3) Either

3.22.7 Do you often eat vegetables in the last year?
(1) Yes, average _____ kg/week (2) No

3.22.8 Do you often eat fruit in the last year?
(1) Yes, average _____ kg/week (2) No

3.22.9 How often do you not eat at home in the last year?
(1) Yes, probably eat outside _____ meals/month (2) No

3.22.10 Usually, does your home use a refrigerator to preserve food?
(1) Yes, probably ____ years (2) No (3) Don't know

3.23 Do you usually drink tea every day?
(1) Yes (2) No (go to question 4.1)

3.24 Which kind of tea do you drink regularly? (Multiple options available)
(1) Green tea (2) Black tea (3) Flower tea (4) Oolong tea (5) Others: _____

4. Health service-related factors

4.1 Have you ever had a hepatitis vaccine? (Multiple choice)
(1) Type A, _____ years (2) Type B, _____years (3) Others, _____ years
(4) No (5) Don't know

4.2 Have you ever had a flu shot?
(1) Yes, _____ years (2) No (3) Don't know

4.3 Did you go to the hospital during the last 12 months?
(1) Yes (2) No (go to question 4.6)

4.3.1 You visited the hospital in the last 12 months __ times, generally in the _____ hospital outpatient service, medical expenses total _____ yuan, other expenses total _____ yuan.

4.4 Have you been to a hospital in the last two weeks?
(1) Yes (2) No (go to question 4.5)

4.4.1 You visited the hospital for ___ visits in the last two weeks. The medical expenses total _____ yuan, and other expenses total _____ yuan.

4.4.2 Your last hospital outpatient registration fee _____,
_____, _____ inspection fee, _____,
and _____ other fees.

4.5 Have you been hospitalized in the last 12 months?
(1) Yes (2) No (go to question 4.6)

4.5.1 What is the reason for hospitalization? _____

4.5.2 You have stayed in ___ hospitals in the past 12 months and lived for a total of ___ days. The hospitalized medical expenses totaled _____ yuan, and other expenses totaled _____ yuan.

4.5.3 Your latest inpatient ward fee is ___yuan, drug charge ___yuan, inspection fee _____ yuan, treatment fee _____ yuan, red envelope _____ yuan, and other cost _____ yuan.

4.6 Have you felt sick in the last two weeks?
(1) Yes (2) No (3) Don't know

4.7 In general, if you are sick, will you go to the hospital?
(1) Yes, as long as I feel uncomfortable (go to question 5.1)
(2) Yes, only if I feel more severe (3) No

4.7.1 Why don't you go to see a doctor (or go when you are serious)? (Multiple choice)
(1) The doctor is too expensive to look down on
(2) The hospital is too far away and it is inconvenient
(3) It is the same to buy medicine for self
(4) It does not matter (5) Other: _____

5. Family related

5.1 Your father's name _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.2 Your mother's name _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3 Do you have siblings?
(1) Yes, brother __, brother __, sister __, sister__ (2) no (go to question 5.4)

5.3.1 Name _____, is your _____, live in local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.2 Name _____, is your _____, live in local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.3 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.4 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.5 Name _____, is your _____, live in local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.6 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.7 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.3.8 Name _____, is your _____, live in local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4 Do you have children?
(1) Yes, son ____, daughter ____ (2) No (go to question 6.1)

5.4.1 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.2 Name _____, is your _____, live in local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.3 Name _____, is your _____, live in the local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.4 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.5 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.6 Name _____, is your _____, live locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.7 Name _____, is your _____, live in local:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

5.4.8 Name _____, which is your _____, lives locally:
(1) Yes, Address: _____ (2) No
Health: (1) Yes (2) No

6. Medical examination related (doctor fills in)
6.1 Height: Exactly _____cm

6.2 Weight: _____ kg

6.3 Waist: Exactly _____cm

6.4 Hip circumference: _____ cm

6.5 Blood pressure _____/_____mmHg (above the on-site physical examination)

6.6 Fasting venous blood glucose value: _____ mmol/L

6.7 HBsAg:
(1) Positive (2) Negative

6.8 Blood Routine: _____

6.9 Stool routine: _____

Investigator: _____

Survey date: _____year _____month ___day

Paper A2.3. Letter of Permission for Data Source Access

復旦大學 公共卫生学院
School of Public Health Fudan University
上海市东安路130号 邮编 200032 130 Dong An Road, Shanghai 200032, P.R.China http//sph.fudan.edu.cn Email: fdsph@fudan.edu.cn

July 23, 2018

To whom it may concern,

This letter is to certify that Miss Cindy Yu has been granted permission to access to the data collected for the project "Rural Deqing Cohort Study". My team and I collected the data from 2006-2014 to develop a risk score model to predict type 2 diabetes for rural Chinese adults. The data was collected with the approval of the Institutional Review Board (IRB) of the Fudan University School of Public Health. There are no personal identifiers present in the available dataset. The written copies of the data were eliminated 5 years after initiation of data collection, and remain in electronic format.

Miss Cindy Yu will use this data for secondary use in her thesis project titled "Effects of ambient fine particulate matter ($PM_{2.5}$) on the incidence of type 2 diabetes mellitus (T2DM) in China – a cohort study". I, Chaowei Fu, was the Principal Investigator for the original collection of the data, and have the authority to grant permission.

Permission granted by

Chaowei Fu 朝伟付, PhD

Professor

Chair

Department of Social Medicine, School of Public Health

Fudan University