International Journal of Epidemiology, 2024, **53(5)**, dyae112
https://doi.org/10.1093/ije/dyae112
Cohort Profile

Cohort Profile

Cohort Profile: The Pearl River Cohort Study

Ying Wang,[1,2] Zhicheng Du,[1,2] Wangjian Zhang,[1,2] Xiaowen Wang,[3,4] Xiao Lin,[1,2] Yu Liu,[5] Yu Deng,[6] Dingmei Zhang,[1] Jing Gu,[1,2] Lin Xu[1,7] and Yuantao Hao[3,4,*]

[1]School of Public Health, Sun Yat-sen University, Guangzhou, China, [2]Research Center for Health Information & Sun Yat-sen Global Health Institute, Sun Yat-sen University, Guangzhou, China, [3]Peking University Center for Public Health and Epidemic Preparedness & Response, Peking University, Peking, China, [4]Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Peking, China, [5]School of Public Health and Management, Guangzhou University of Chinese Medicine, Guangzhou, China, [6]Laboratory Animal Center, Sun Yat-sen University, Shenzhen, China and [7]School of Public Health, The University of Hong Kong, Hong Kong, China

*Corresponding author. Peking University Center for Public Health and Epidemic Preparedness & Response, Peking University, No. 38 Xueyuan Road, Haidian District, Peking 100191, China. E-mail: haoyt@bjmu.edu.cn

Keywords: Infectious disease, non-communicable diseases, Cohort Study, China.

Key Features

- The Pearl River Cohort Study (PRCS) is a prospective, community-based study conducted in the Pearl River Delta Region of China, aiming to explore the epidemiological features of predominant infectious diseases, particularly hepatitis B virus (HBV) and human immunodeficiency virus (HIV), and examine the corresponding trends and determinants. PRCS further assesses the burden of non-communicable diseases and explores the underlying risk factors, including genetic and environmental factors.
- From 2009 to 2015, 1 012 313 participants (477 829 men and 534 484 women) were recruited from Guangzhou, Zhongshan and Shenzhen in the Pearl River Delta Region. The study assessed sociodemographic and behaviour factors and collected blood samples. HBV and HIV screenings were also conducted to determine the prevalence of these infections.
- A second wave of follow-up was carried out between 2018 and 2020, with 10% of the participants being re-surveyed following baseline survey procedures. Participants' identifiers are securely linked to their local medical records for thorough follow-up, enabling real-time tracking of major health outcomes without direct periodic contact.
- Blood biomarkers, including inflammation markers and genome-wide DNA methylation profiles, provide valuable insights into the underlying biological mechanisms.
- The PRCS welcomes specific queries, proposals for collaboration and reasonable requests for data access. Please direct such inquiries to the corresponding author at [haoyt@bjmu.edu.cn].

Why was the cohort set up?

Infectious diseases, notably hepatitis B virus (HBV) and human immunodeficiency virus (HIV) infection, represent a growing public health concern in China.[1] In 2019, China was responsible for approximately one-third of the global HBV infections (296 million).[1,2] Comparatively, about 39.0 million individuals were living with HIV globally by 2022.[3] The prevalence of these infectious diseases varies across countries. In 2019, the Western Pacific region showed an HBV prevalence of 7.1%, with a prevalence of 7.8% in China.[4] By 2020, China was reported with 1.053 million HIV cases and 0.35 million related deaths.[5] Guangdong, an economically prosperous province in South China, recorded an 8.76% prevalence of hepatitis B surface antigen (HBsAg) among the general population between 2014 and 2015.[6] This high prevalence was probably attributed to factors such as population density, migration patterns, or limited health resources.[6] Furthermore, the province recorded 78 200 HIV cases by 2021, indicating the urgent need for intensified public health interventions.[7]

In 2009, the Community-based Collaborative Innovation (CCI) project,[8] funded by the National Science and Technology Major Project of China was launched. It was recognized as a Major Special Project in the National Medium- and Long-term Program for Science and Technology Development, aimed at the prevention and control of major infectious diseases, including HBV and HIV. The Pearl River Cohort Study (PRCS) was initiated in Guangdong province as a part of the CCI project. As a community-based cohort, PRCS collected demographic and health-related data, as well

Received: 15 October 2023. **Editorial Decision:** 25 June 2024. **Accepted:** 23 August 2024

as comprehensive information on chronic diseases, to facilitate a deeper understanding of the health landscape in the community and provide valuable insights for targeted interventions. By leveraging such insights, community health programs can be tailored to address specific health needs and challenges and improve the quality of life.[8,9]

Additionally, in 2019, non-communicable diseases (NCDs), such as cardiovascular disease (CVD), cancer, chronic respiratory diseases and diabetes, were responsible for 88.5% of all deaths in China.[10] The prevalence of NCDs in China has been rising, potentially due to socio-economic development, advancements in health services, increased average life expectancy and a confluence of risk factors, including aging and adverse health behaviours.[11–13] For example, the number of individuals with diabetes in China grew from 90 million in 2011 to 140.9 million in 2021.[14,15] Furthermore, China accounted for a substantial proportion of CVD cases worldwide, with an estimated 330 million cases in 2019.[16]

The PRCS primarily aims to elucidate the epidemiological characteristics of infectious diseases, specifically HBV and HIV, and to explore their trends and determinants. As a community-based cohort, the PRCS is well-equipped to continuously monitor dynamic changes in risk factors (including genetics and environmental factors), disease trajectories, and the burden of both infectious diseases and NCDs. Furthermore, the PRCS offers insights into the biological mechanisms linking exposures with health outcomes, facilitated by its comprehensive biobanking.

Who is in the cohort?

The PRCS was initiated in the Pearl River Delta Region, with Guangzhou, Zhongshan and Shenzhen selected as the representative areas according to geographic and socio-economic development (Figure 1). The baseline investigation spanned from 2009 to 2015. Inclusion criteria included being permanent residents, capable of undergoing a physical examination, and willing to participate with a signed informed consent form. Exclusion criteria included: (i) an inability to complete the questionnaire, undergo blood sampling or take physical examination; (ii) the presence of mental or cognitive abnormalities. As a community-based cohort, participants were recruited using a random cluster sampling technique, with communities in these three cities serving as the sampling units. The health committees of each region facilitated this process by issuing official documents to community health service centres to encourage residents to participate. Potential participants were identified using official residential records, ensuring a wide and diverse representation. During the recruitment, approximately 2.8 million individuals who participated in Basic Public Health Services were approached, with a final enrolment of 1 012 313 participants, representing a participation rate of about 36%. To assess the representativeness of our cohort, we compared demographic characteristics such as age and gender with those of the general population of Guangdong Province, as reported in the Seventh Population Census. Before enrolment, residents were comprehensively informed about the study and provided written informed consent, permitting access to their medical records and the extended storage of their blood for anonymized research purposes.

A total of 1 012 313 participants were recruited, including 190 253 from Zhongshan, 167 945 from Shenzhen and 654 115 from Guangzhou. Table 1 shows that 47.2% of the participants were men, 98.2% identified as Han ethnicity, and the median age was 39.63 years [interquartile range (IQR) = 28.48]. Our cohort displays a slightly higher proportion of individuals over 60 years old (+5.55%) and females (+5.87%) compared with the general population of Guangdong Province, as reported in the Seventh Population Census Data Report. In addition, 69.5% of the participants were married, and most participants (82.6%) completed a minimum of 6 years of schooling (middle school or higher). The proportion of men who were current smokers (7.5% vs 0.3%) and ever drinkers (6.6% vs 1.2%) was greater than their counterparts among women in the group over 18 years old. The average body mass index (BMI) was 21.87 [standard deviation (SD) = 2.83] kg/m^2 for men and 21.48 (SD = 2.99) kg/m^2 for women, and the prevalence of hypertension was 50.3% in men and 47.8% in women.

The mean and SD of liver function and routine blood biomarkers on 114 728 participants are shown in Table 2. Survival outcomes, determined through linkages with the Death Registry, operated by the Guangzhou and Zhongshan Center for Disease Control and Prevention (CDC) till 31 December 2020, are shown in Table 3.

How often have the respondents been followed up?

Long-term follow-ups are carried out through both active and passive approaches. The reassessment, rooted in community-based sampling, not only reiterated baseline procedures but also introduced additional survey items, such as those about sleep quality and dietary habits. Blood samples were also collected. For active follow-up, after the initial survey, we re-evaluated a subset of participants every 3–5 years following consistent procedures. During the first follow-up from 2018 to 2020, constrained by funding limitations, we reached out to a substantial subset of the original cohort, comprising approximately 30% of the baseline participants. This outreach was conducted through text messages and phone calls. Of those invited, around 10% of the total initial cohort, which equates to approximately 126 328 participants, responded and agreed to participate in the follow-up survey. This active follow-up is facilitated through our close collaboration with community healthcare teams, including community doctors and general practitioners, and in conjunction with the Basic Public Health Services provided to the permanent residents. The percentage of missing data for variables at baseline and first follow-up is shown in Supplementary Table S1 (available as Supplementary data at *IJE* online).

Regarding passive follow-up, all participants are continuously monitored for health service usage and cause-specific mortality. This process is streamlined by integrating the unique medical insurance card and national ID, allowing for comprehensive tracking without direct contact. The areas encompassed by our study are integrated within the framework of national Disease Surveillance Points system, which is instrumental in generating mortality statistics nationwide. We ascertain the vital status of participants through examination of the official residential records and the procurement of death certificates, as reported to the regional CDC. Future follow-ups for disease-specific hospital admissions will be

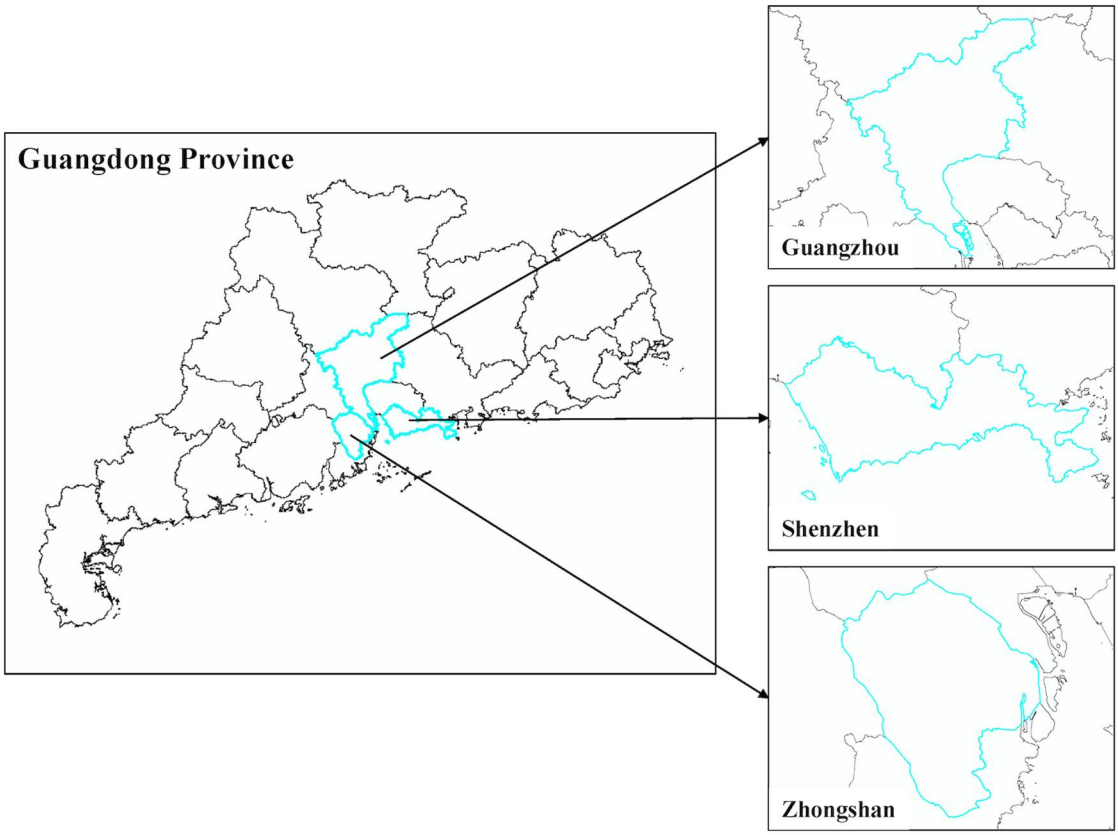

Figure 1. Geographical representation of the three survey sites within Guangdong Province, China

primarily based on electronic linkage with the national health insurance claims databases. Follow-up data for younger adults with hypertension and diabetes is gathered through linkage with healthcare and new national health insurance databases, facilitating systematic monitoring of medical visits, medication and treatment regimens, and enabling the evaluation of patient prognosis and complication rates. Follow-up data for the elderly with hypertension and diabetes are obtained by linking to the National Basic Public Health dataset, enabling dynamic health monitoring.

What has been measured?

A computer-based standardized structured questionnaire was used to collect detailed information through face-to-face interviews. Information collected included demographic characteristics, behaviour factors, occupational exposures, and family and personal history of diseases [i.e. HBV infection, hypertension, diabetes, coronary heart disease, chronic obstructive pulmonary disease (COPD), cancer and other diseases]. The awareness rates related to HBV and acquired immune deficiency syndrome (AIDS) knowledge were also assessed (Table 4). Physical measurements, including height, weight, waist circumference, blood pressure and pulse rate were taken using standardized instruments and protocols by well-trained nurses. Upon completion of the assessment, research assistants reviewed the entries for completeness and accuracy, ensuring that data were consistently coded according to unified guidelines. The types of variables are shown in Supplementary Table S2 (available as Supplementary data at *IJE* online).

For each participant, a 5-mL fasting venous blood sample was collected on the same date as the baseline examination. Participants were instructed to fast overnight before the examination. Blood was collected in EDTA anticoagulant and coagulation-promoting tubes after the information of participants was verified. The serum and coagulum collected in coagulation-promoting tubes, as well as the plasma and blood cells collected in EDTA tubes, are all stored at −80°C. The details of the blood processing methods can be found in the Supplementary material available at *IJE* online. The specimens were transported on dry ice to the Biobank at the School of Public Health, Sun Yat-Sen University, for cryopreservation at −80°C. Lipid profiles and liver function markers were analyzed using enzymatic assays with an automatic analyzer, and the serum specimens were screened for hepatitis B-related biomarkers using enzyme-linked immunosorbent assay test kits.

What has been found?

Up to December 2023, 32 studies had been published using the PRCS data, which investigated the incidence of infectious diseases and explored risk factors for both infectious diseases and NCDs. Our study showed an HBsAg prevalence of 8.76% (8.82% for males and 8.65% for females),[6] underscoring the substantial strides made in hepatitis B control in Guangdong. This represents a notable decline from the 16.67% prevalence observed in 1992[17] and 11.10% in 2006.[18] Notably, the highest prevalence was recorded among individuals aged 35–39 years—with a prevalence of 17.71% for the whole group, 25.58% among males and 16.27% among females. Additionally, we employed two variable

Table 1. Baseline characteristics of the Pearl River Cohort Study participants

Characteristics	Men	Women	Total
Number of participants	477 829	534 484	1 012 313
Age (year)			
<18	52 667 (11.0%)	47 407 (8.9%)	100 074 (9.9%)
18 ≤ 30	98 359 (20.6%)	136 193 (25.5%)	234 552 (23.2%)
30 ≤ 40	82 688 (17.3%)	95 445 (17.9%)	178 133 (17.6%)
40 ≤ 50	84 969 (17.8%)	84 122 (15.7%)	169 091 (16.7%)
50 ≤ 60	73 769 (15.4%)	75 129 (14.0%)	148 898 (14.7%)
60 ≤ 70	45 146 (9.5%)	46 045 (8.6%)	91 191 (9.0%)
70 ≤ 80	27 339 (5.7%)	33 256 (6.2%)	60 595 (6.0%)
≥80	12 892 (2.7%)	16 887 (3.2%)	29 779 (2.9%)
Median (IQR)	40.59 (28.44)	38.65 (28.48)	39.63 (28.48)
Ethnicity			
Han	465 238 (98.1%)	490 771 (98.3%)	956 009 (98.2%)
Minority	8928 (1.9%)	8593 (1.7%)	17 521 (1.8%)
Marital status			
Never married	135 563 (32.0%)	116 710 (23.8%)	252 273 (27.6%)
Married	280 642 (66.3%)	354 168 (72.1%)	634 810 (69.4%)
Widowed	3627 (0.9%)	15 201 (3.1%)	18 828 (2.1%)
Divorced	3374 (0.8%)	4700 (1.0%)	8074 (0.9%)
Education			
No formal school	5917 (1.7%)	9137 (2.6%)	15 054 (2.1%)
Primary school	48 273 (14.0%)	59 275 (16.5%)	107 548 (15.3%)
Middle school	105 170 (30.4%)	101 435 (28.3%)	206 605 (29.3%)
High school	121 890 (35.3%)	120 663 (33.7%)	242 553 (34.5%)
College/university or higher	64 447 (18.6%)	67 849 (18.9%)	132 296 (18.8%)
Smoking status			
Never	403 096 (91.9%)	452 126 (99.5%)	855 222 (95.7%)
Ever smoker	6148 (1.4%)	844 (0.2%)	6992 (0.8%)
Current smoker	29 569 (6.7%)	1410 (0.3%)	30 979 (3.5%)
Alcohol drinking			
Never	374 976 (94.1%)	401 097 (98.9%)	776 073 (96.5%)
Ever drinker	23 367 (5.9%)	4478 (1.1%)	27 845 (3.5%)
Exercise frequency			
Low	318 905 (78.1%)	321 063 (76.6%)	639 968 (77.4%)
Moderate	43 723 (10.7%)	42 230 (10.1%)	85 953 (10.4%)
High	45 548 (11.2%)	55 613 (13.3%)	101 161 (12.2%)
BMI (kg/m^2)			
<18.5	29 251 (9.3%)	40 870 (11.8%)	70 121 (10.6%)
18.5 ≤ 24	229 594 (73.3%)	248 585 (71.9%)	478 179 (72.5%)
24 ≤ 28	48 330 (15.4%)	47 630 (13.8%)	95 960 (14.6%)
≥28	6169 (2.0%)	8789 (2.5%)	14 958 (2.3%)
Mean (SD)	21.87 (2.83)	21.48 (2.99)	21.67(2.92)
Height (cm)			
<155	30 691 (9.5%)	106 747 (30.0%)	137 438 (20.3%)
155 ≤ 160	10 503 (3.3%)	100 128 (28.1%)	110 631 (16.3%)
160 ≤ 165	32 740 (10.2%)	105 218 (29.6%)	137 958 (20.4%)
≥165	247 130 (77.0%)	43 929 (12.3%)	291 059 (43.0%)
Mean (SD)	164.97 (15.59)	156.11 (12.31)	160.31 (14.65)
Hypertension			
Yes	154 201 (50.3%)	164 441 (47.8%)	318 642 (49.0%)
No	152 219 (49.7%)	179 558 (52.2%)	331 777 (51.0%)

IQR, interquartile range; SD, standard deviation; BMI, body mass index.

selection algorithms, the two-stage hybrid and bootstrap ranking procedures, in a large-scale epidemiology survey of HBV infection, which identified three important factors associated with HBV infection—the individual history of the hepatitis B vaccination, as well as family and individual infection history.[19]

Using data from the PRCS, we also found that long-term exposure to particulate matter (PM) was positively associated with hypertension,[20] diabetes,[20] and mortality risks from CVD,[21] cerebrovascular diseases[22] and COPD.[23] During an average follow-up of 8.0 years, with 4 753 965 person-years of 580 757 participants in Guangzhou, 37 578 deaths occurred. This included 14 448 from CVD, 7260 from respiratory diseases and 15 870 from other causes. We analyzed these associations between time-varying PM concentrations and all-cause mortality using a linear model, with each $1\,\mu g/m^3$ increase in PM levels serving as the independent variable. Higher PM concentrations were associated with higher risk of all-cause mortality, with hazard ratios (HRs) of 1.042 [95% confidence interval (CI): 1.037–1.046), 1.031 (95% CI: 1.028–1.033) and 1.029 (95% CI: 1.027–1.031] per $1\,\mu g/m^3$ increase in PM_1 (PM with an aerodynamic diameter-$\leq 1\,\mu m$), $PM_{2.5}$ (PM with an aerodynamic diameter $\leq 2.5\,\mu m$) and PM_{10} (PM with an aerodynamic diameter $\leq 10\,\mu m$) concentrations, respectively.[24] Furthermore, consistent with findings from previous studies, we observed that participants with unhealthy lifestyle factors, such as excessive alcohol consumption or insufficient sleep, were more susceptible to the adverse effects of PM exposure.[25]

Table 2. Baseline characteristics of liver function and routine blood biomarkers in 114 728 participants of the Pearl River Cohort Study in 2009–2015

Variable	Total ($N = 114\,728$)	Men ($N = 42\,888$)	Women ($N = 71\,840$)
Liver function, mean (SD)			
Albumin (g/L)	46.38 (3.41)	46.74 (3.53)	46.17 (3.32)
Alanine aminotransferase (U/L)	20.96 (17.69)	22.81 (20.22)	19.86 (15.89)
Aspartate aminotransferase (U/L)	24.51 (12.59)	26.02 (14.82)	23.61 (10.95)
Direct bilirubin (μmol/L)	3.27 (1.51)	3.53 (1.76)	3.12 (1.30)
Total bilirubin (μmol/L)	10.84 (4.58)	11.14 (5.14)	10.66 (4.21)
Blood routine, mean (SD)			
White blood cell count (10^9/L)	6.46 (1.83)	6.79 (2.00)	6.26 (1.69)
Percent of monocytes (%)	4.70% (1.82%)	4.95 (1.84%)	4.55 (1.79%)
Monocyte count (10^9/L)	0.30 (0.14)	0.33 (0.15)	0.28 (0.13)
Red cell volume distribution width-variable coefficient (%)	14.39% (1.41%)	14.39% (1.34%)	14.39% (1.44%)
Red cell volume distribution width-standard deviation (fL)	53.63 (7.27)	53.67 (7.65)	53.60 (7.03)
Red blood cell count (10^{12}/L)	4.61 (0.53)	4.82 (0.54)	4.49 (0.48)
Haematocrit (%)	45.33% (5.03%)	47.41% (5.40%)	44.08% (4.35%)
Lymphocyte percentage (%)	38.26% (9.66%)	38.10% (9.97%)	38.36% (9.48%)
Lymphocyte count (10^9/L)	2.43 (0.80)	2.54 (0.87)	2.36 (0.75)
Mean corpuscular volume (fL)	98.92 (10.62)	98.91 (10.91)	98.93 (10.45)
Mean red blood cell haemoglobin content (pg)	29.34 (3.22)	29.58 (3.21)	29.19 (3.22)
Mean corpuscular haemoglobin concentration (g/L)	297.32 (21.52)	299.92 (20.60)	295.77 (21.91)
Mean platelet volume (fL)	9.02 (0.95)	8.94 (0.94)	9.06 (0.95)
Percent of basophilic granulocyte (%)	0.59% (0.31%)	0.61% (0.32%)	0.58% (0.31%)
Basophilic granulocyte count (10^9/L)	0.04 (0.02)	0.04 (0.03)	0.04 (0.02)
Percentage of eosinophilic granulocyte (%)	3.19% (2.35%)	3.73% (2.67%)	2.87% (2.07%)
Eosinophil count (10^9/L)	0.21 (0.17)	0.25 (0.20)	0.18 (0.14)
Haemoglobin (g/L)	134.33 (13.98)	141.65 (14.42)	129.95 (11.7)
Platelet count (10^9/L)	262.82 (69.80)	260.3 (73.70)	264.33 (67.32)
Plateletcrit (%)	0.23% (0.06%)	0.23% (0.06%)	0.24% (0.06%)
Percent of neutrophile granulocyte (%)	53.26% (9.87%)	52.61% (10.10%)	53.64% (9.71%)
Neutrophil count (10^9/L)	3.49 (1.40)	3.62 (1.54)	3.41 (1.31)

SD, standard deviation.

Table 3. Survival outcomes of Pearl River Cohort Study participants in Guangzhou and Zhongshan followed until 31 December 2020

Outcomes	Total	Men	Women
All-cause death	42 285	23 517	18 768
Cardiovascular diseases	15 099 (35.71%)	7532 (32.03%)	7567 (40.32%)
Ischaemic heart disease	7009 (16.58%)	3537 (15.04%)	3472 (18.50%)
Stroke	5359 (12.67%)	2716 (11.55%)	2643 (14.08%)
Respiratory diseases	7579 (17.92%)	4358 (18.53%)	3221 (17.16%)
Lower respiratory infections	3728 (8.82%)	1983 (8.43%)	1745 (9.30%)
Chronic obstructive pulmonary disease	2346 (5.55%)	1589 (6.76%)	757 (4.03%)
Lung cancer	3323 (7.86%)	2199 (9.35%)	1124 (5.99%)

Based on the PRCS, we also observed a protective effect of green space against respiratory and lung cancer mortality.[26,27] Specifically, each IQR increment in greenness exposure was associated with a lower risk of respiratory and lung cancer mortality, with an HR of 0.88 (95% CI: 0.84, 0.91)[26] and 0.89 (95% CI: 0.83, 0.96),[27] respectively. The associations could be mediated through various pollutants and factors, including $PM_{2.5}$, PM_{10}, nitrogen dioxide, temperature and physical activity.

Our findings have also enriched specific research in public health, machine learning and causal inference, contributing valuable insights into the epidemiology of HBV and the use of advanced analytical methods.[19,28–30]

What are the main strengths and limitations?

To the best of our knowledge, this study represents one of the most expansive community-based studies in South China, aiming to explore the associations of diverse environmental exposures and genetic risk factors with the incidence of infectious diseases, NCDs and specific-cause mortality. The distinctive characteristics of the Pearl River Delta region offer the opportunity for this large cohort study to make unique contributions to the existing literature. For instance, an in-depth exploration of the Cantonese culinary tradition could yield insights into its potential health implications. Additionally, the region's rapid industrialization prompts a crucial examination of the health impacts of pollution, climate change and other environmental concerns. With prominent tech hubs present in the region, there is also a compelling need to investigate technology's role in health, from the advancements of telemedicine to the broader effects of the digital economy on mental and physical health. In addition, our study collected comprehensive information at baseline and during subsequent follow-ups. Moreover, we secured biological specimens for laboratory testing to obtain information on lipid profiles, liver function and genetic variants. This information, along with the burden of disease

Table 4. Baseline data collation overview in the Pearl River Cohort Study

Category	Variable/exposure
Baseline questionnaire	
Demographics	Age, sex, workplace, ethnicity, self-reported blood type, education, occupation, marital status, medical insurance
Personal health behaviour	Smoking status, alcohol consumption, exercise (frequency of taking exercise within a certain period of time), migratory status, dietary habits
Environment exposure	Kitchen facility, cooking fuels, sources of drinking water, occupational exposures
Family and personal disease histories	History of hypertension, diabetes, coronary heart disease, COPD, cancer, stroke, mental disorder, tuberculosis, hepatitis, AIDS
Medical history and wellbeing	Personal history of surgery, transfusion or trauma, family history of HBV infection, history of hepatitis B vaccination, disability status, self-reported general health, self-assessment of activities of daily living, cognitive function, depressive symptoms
Inpatient situation	History of hospitalization, cause of inpatient admission, inpatient admission date, healthcare institution
Medication management	Diagnosed medical conditions, use of medicines, time of medication use, medication adherence
HBV and AIDS/HIV knowledge	HBV knowledge (contagiousness, transmission route, vaccination timing for newborns, prevention methods, and treatment) AIDS/HIV knowledge (acquisition of information, transmission routes, measures to reduce spread)
Anthropometric and vital measurements	
	Height, weight, waist circumference, pulse rate, blood pressure
Laboratory test	
Blood routine	White blood cell count, red blood cell, platelet, haemoglobin, neutrophil count, lymphocyte count, mean corpuscular volume, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, white blood cell differential count, red cell volume distribution width, platelet haematocrit
Liver function	Alanine aminotransferase, aspartate transaminase, albumin, total bilirubin, conjugated bilirubin
Blood lipids	Total cholesterol, triglycerides, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol
Liver ultrasound	B-mode ultrasound (Kunlun Resona I9.)
Serological screening	Hepatitis B surface antigen, hepatitis B surface antibody, human immunodeficiency virus antibody

COPD, chronic obstructive pulmonary disease; HBV, hepatitis B virus; AIDS, acquired immunodeficiency syndrome.

evaluation framework and tools developed in our previous studies, underpins evidence-based policymaking and community resource allocation. Notably, we achieved seamless integration with the routine disease surveillance system and the CDC's Death Registry, ensuring accurate disease incidence and mortality data capture, of multiple diseases, complemented by triennial to quinquennial exposure updates via re-interviews.

However, some limitations merit acknowledgment. Firstly, as the cohort was established in Guangdong province, it is not nationally representative, which could potentially limit the generalizability of our findings. Additionally, similar to most previous prospective cohort studies, selection bias may be possible in this cohort. Although participants in a long-term cohort might be more health-conscious and likely to have a healthier lifestyle than others, our provision of complementary healthcare services might have bolstered participation among lower socioeconomic strata. Furthermore, in our study, we found that the prevalence of alcohol consumption among men in Southern China was lower than that reported in previous studies, which predominantly focused on Northern Chinese populations.[31,32] This discrepancy might be attributed to cultural differences between the North and South, as alcohol consumption is generally less common in the South. Moreover, potential volunteer bias might also play a role. PRCS participants may be somewhat healthier than the broader Southern Chinese population, likely because they lead healthier lifestyles. However, it is worth noting that the prevalence of major chronic diseases like diabetes and hypertension[20] in our sample was consistent with national data,[32,33] so representativeness should not be a major concern.

Can I get hold of the data? Where can I find out more?

In adherence to rigorous ethical standards and to safeguard the confidentiality of participant information, access to the PRCS cohort data and biospecimens is currently restricted to our research team. However, we are committed to promoting epidemiological research through collaboration. We welcome inquiries and proposals from researchers interested in using our data for complementary studies. Please direct detailed requests for collaboration or data access to Prof. Yuantao Hao via email at haoyt@bjmu.edu.cn.

Ethics approval

The study was approved by the Human Ethics Committee at Sun Yat-sen University (No. L2017030).

Data availability

See 'Can I get hold of the data? Where can I find out more?' above.

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

Y.W. and L.X. helped conceive the design of the study, analyzed the data and draft the manuscript; Y.H. helped conceive the design of the study and revised the article for critically

important intellectual content; Z.D., W.Z., Y.D., D.Z. and J. G. performed data collection; X.L., Y.L. and X.W. helped draft the manuscript; Z.D. and W.Z. supervised data analyses. All authors have read and approved the final version of the manuscript, and agree with the order of the presentation of the authors.

Funding

This work was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China (2012ZX10004902, 2012ZX10005009, 2013ZX10004901 and 2018ZX10715004), National Natural Science Foundation of China (82204162, 82304248, 81973150), Basic and Applied Basic Research Foundation of Guangdong Province (2022A1515010823, 2022A1515110263). Professor Yuantao Hao gratefully acknowledges the support of K.C. Wong Education Foundation from Peking University.

Acknowledgements

The most important acknowledgment is to all the participants and the staff of the survey teams in each of the study sites and communities. Thanks for the cooperation of Sun Yat-sen university, Guangdong Provincial Center for Disease Control and Prevention, Shenzhen Center for Disease Control and Prevention, Zhongshan Center for Disease Control and Prevention, The Third affiliated Hospital, Sun Yat-sen University, The Eighth People's Hospital of Guangzhou City, Guangzhou Center for Disease Control and Prevention, Guangzhou Yuexiu District Center for Disease Control and Prevention, Guangzhou Liwan District Center for Disease Control and Prevention, Shenzhen Luohu District Center for Disease Control and Prevention and Shenzhen Baoan District Center for Disease Control and Prevention.

Conflict of interest

None declared.

References

1. World Health Organization. *Global Progress Report on HIV, Viral Hepatitis and Sexually Transmitted Infections*. 2021. https://www.who.int/publications/i/item/9789240027077 (21 August 2023, date last accessed).
2. Chen S, Li J, Wang D, Fung H, Wong LY, Zhao L. The hepatitis B epidemic in China should receive more attention. *Lancet* 2018; **391**:1572.
3. World Health Organization. *HIV and AIDS*. 2023. https://www.who.int/news-room/fact-sheets/detail/hiv-aids? (21 August 2023, date last accessed).
4. GBD 2019 Hepatitis B Collaborators. Global, regional, and national burden of hepatitis B, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Gastroenterol Hepatol* 2022;7:796–829.
5. He N. Research progress in the epidemiology of HIV/AIDS in China. *China CDC Wkly* 2021;**3**:1022–30.
6. Zeng F, Guo P, Huang Y *et al.* Epidemiology of hepatitis B virus infection: results from a community-based study of 0.15 million residents in South China. *Sci Rep* 2016;**6**:36186.
7. Health Commission of Guangdong Province. *The AIDS Epidemic in Guangdong Is Generally at a Low Prevalence Level.* http://wsjkw.gd.gov.cn/banner/content/post_3691091.html (21 August 2023, date last accessed).
8. Ruan B, Yu Z, Yang S *et al.* Establishment and development of national community-based collaborative innovation demonstration areas to achieve the control target of hepatitis B in China. *BMC Infect Dis* 2019;**19**:617.
9. Vandiver T, Anderson T, Boston B, Bowers C, Hall N. Community-based home health programs and chronic disease: synthesis of the literature. *Prof Case Manag* 2018;**23**:25–31.
10. Peng W, Chen S, Chen X, *et al.* Trends in major non-communicable diseases and related risk factors in China 2002-2019: an analysis of nationally representative survey data. *Lancet Reg Health West Pac* 2023;**43**:100809.
11. Yang G, Wang Y, Zeng Y *et al.* Rapid health transition in China, 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* 2013;**381**:1987–2015.
12. GBD 2017 Population and Fertility Collaborators. Population and fertility by age and sex for 195 countries and territories, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;**392**:1995–2051.
13. Yin P, Brauer M, Cohen AJ *et al.* The effect of air pollution on deaths, disease burden, and life expectancy across China and its provinces, 1990–2017: an analysis for the Global Burden of Disease Study 2017. *Lancet Planet Health* 2020;**4**:e386–98.
14. Lixin G. 2021 annual review of major progress in diabetes. *Zhong Hua Tang Niao Bing Za Zhi* 2022;**14**:1–8.
15. International Diabetes Federation. *IDF Diabetes Atlas*, 10th edn. Brussels, Belgium, 2021. https://www.diabetesatlas.org (16 August 2024, data last accessed).
16. Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on cardiovascular health and diseases in China 2021: an updated summary. *Biomed Environ Sci* 2022; **35**:573–603.
17. Xia G, Liu C, Cao H *et al.* Prevalence of hepatitis B and C virus infections in the general Chinese population. results from a nationwide cross-sectional seroepidemiologic study of hepatitis A, B, C, D, and E virus infections in China. *Int Hepatol Commun* 1992; **5**:62–73.
18. Xiao J, Zhang J, Wu C *et al.* Impact of hepatitis B vaccination among children in Guangdong Province, China. *Int J Infect Dis* 2012;**16**:e692–96.
19. Guo P, Zeng F, Hu X *et al.* Improved variable selection algorithm using a LASSO-type penalty, with an application to assessing hepatitis B infection relevant factors in community residents. *PLoS One* 2015;**10**:e134151.
20. Cai H, Du Z, Lin X *et al.* Interactions between long-term ambient particle exposures and lifestyle on the prevalence of hypertension and diabetes: insight from a large community-based survey. *J Epidemiol Community Health* 2023;**77**:440–46.
21. Zhang Y, Wang Y, Du Z *et al.* Potential causal links between long-term ambient particulate matter exposure and cardiovascular mortality: new evidence from a large community-based cohort in South China. *Ecotoxicol Environ Saf* 2023; **254**:114730.
22. Chen S, Lin X, Du Z *et al.* Potential causal links between long-term ambient particulate matter exposure and cerebrovascular mortality: insights from a large cohort in southern China. *Environ Pollut* 2023;**328**:121336.
23. Wang Y, Du Z, Zhang Y *et al.* Long-term exposure to particulate matter and COPD mortality: insights from causal inference methods based on a large population cohort in southern China. *Sci Total Environ* 2023;**863**:160808.
24. Wang Y, Wei J, Zhang YQ *et al.* Estimating causal links of long-term exposure to particulate matters with all-cause mortality in South China. *Environ Int* 2023;**171**:107726.
25. Ju X, Yimaer W, Du Z *et al.* The impact of monthly air pollution exposure and its interaction with individual factors: insight from a large cohort study of comprehensive hospitalizations in Guangzhou area. *Front Public Health* 2023; **11**:1137196.

26. Wu W, Du Z, Wang Y *et al.* The complex role of air pollution on the association between greenness and respiratory mortality: insight from a large cohort, 2009–2020. *Sci Total Environ* 2023; **899**:165588.
27. Wu W, Wu G, Wei J *et al.* Potential causal links and mediation pathway between urban greenness and lung cancer mortality: result from a large cohort (2009 to 2020). *Sustain Cities Soc* 2024;**101**:105079.
28. Tian X, Chong Y, Huang Y *et al.* Using machine learning algorithms to predict hepatitis B surface antigen seroclearance. *Comput Math Methods Med* 2019;**2019**:6915850.
29. Huang Y, Li X, Zheng X *et al.* Bayesian network to predict hepatitis B surface antigen seroclearance in chronic hepatitis B patients. *J Viral Hepat* 2020;**27**:1326–37.
30. Wang Y, Du Z, Lawrence WR *et al.* Predicting hepatitis B virus infection based on health examination data of community population. *Int J Environ Res Public Health* 2019;**16**:4842.
31. Zhao X, Hong F, Yin J *et al.*; China Multi-Ethnic Cohort (CMEC) Collaborative Group. Cohort Profile: the China Multi-Ethnic Cohort (CMEC) study. *Int J Epidemiol* 2021;**50**:721–21l.
32. Chen Z, Chen J, Collins R *et al.*; China Kadoorie Biobank (CKB) Collaborative Group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;**40**:1652–66.
33. Wang F, Zhu J, Yao P *et al.* Cohort Profile: the Dongfeng-Tongji cohort study of retired workers. *Int J Epidemiol* 2013; **42**:731–40.

International Journal of Epidemiology, 2024, 53, 1–8
https://doi.org/10.1093/ije/dyae112
Cohort Profile