BMJ Open

Chronic non-communicable diseases: Hainan prospective cohort study

Xingbo Gu,[1] Liuting Lin,[2,3] Chanjuan Zhao,[1] Ling Wu,[1] Yumei Liu,[2] Limin He,[2] Guotian Lin,[2] Yingzi Lin,[2] Fan Zhang [2]

[1]Department of Biostatistics, International School of Public Health and One Health, Hainan Medical University, Haikou, Hainan, China
[2]Laboratory of Tropical Environment and Health, International School of Public Health and One Health, Hainan Medical University, Haikou, Hainan, China
[3]Department of Science and Education, Hainan Cancer Hospital, Haikou, China

Correspondence to
Professor Fan Zhang;
zhangfan@hainmc.edu.cn and
Yingzi Lin;
linyingzi@hainmc.edu.cn

ABSTRACT

Purpose The Hainan Cohort was established to investigate the incidence, morbidity and mortality of non-communicable diseases and their risk factors in the community population.

Participants The baseline investigation of the Hainan Cohort study was initiated in five main areas of Hainan, China, from June 2018 to October 2020. A multistage cluster random-sampling method was used to obtain samples from the general population. Baseline assessments included a questionnaire survey, physical examination, blood and urine sample collection, and laboratory measurements, and outdoor environmental data were obtained.

Findings to data A total of 14 443 participants aged 35–74 years were recruited at baseline, with a participation rate of 90.1%. The mean age of the participants was 48.8 years; 51.8% were men, and 83.7% had a secondary school or higher education. The crude prevalence of diabetes, coronary heart disease, stroke, hypertension, hyperuricaemia, chronic bronchitis, pulmonary tuberculosis, asthma, cancer, chronic hepatitis and metabolic syndrome were 8.6%, 9.2%, 2.0%, 37.1%, 7.1%, 2.3%, 1.4%, 2.1%, 4.1%, 2.2% and 14.5%, respectively.

Future plans The Hainan Cohort is a dynamic cohort with no end date. All participants will be monitored annually for cause-specific mortality and morbidity until death. Long-term follow-up will be conducted every 5 years. The baseline population is considered to expand in the next wave of follow-up, depending on the availability of funding support.

INTRODUCTION

Chronic non-communicable diseases (NCDs), such as cardiovascular diseases, cancer, diabetes and chronic respiratory diseases, are the most important public health problems globally.[1–3] The main characteristics of NCDs are that they are common, long-lasting, and cannot be or are rarely cured completely.[4] According to the WHO, NCDs are the leading cause of death worldwide, accounting for 71% of all deaths annually.[5] In recent years, with the rapid development of China's economy and continuous improvements in people's living standards, traditional living environments, diets and lifestyles have changed, leading to significant changes in

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The Hainan Cohort study is the first independent biospecimen-based prospective study on non-communicable diseases (NCDs) in tropical China.
⇒ The sample size was relatively large, which gives us sufficient power to explore the relationship between risk factors and common chronic NCDs.
⇒ The cohort database linked the medical insurance system, the mortality surveillance system and the national basic public health service platform and will combined these data with those obtained from long-term follow-up surveys to ensure that obtain accurate endpoint information.
⇒ The baseline survey did not collect the data of eligible people who did not participate in the survey, which may lead to selection bias.
⇒ More than 98% of the participants in the cohort were Han Chinese, caution should be used when the results were applied to other populations.

the main risk factors for NCDs in the Chinese population.[6] Although substantial progress in reducing the burden of NCDs in China has been made, the prevalence of NCDs continues to increase; moreover, the burdens imposed by NCD subcategories have variably persisted in different provinces of China.[7–9]

Hainan is a tropical island in the southernmost part of China and is China's second largest island after Taiwan.[10] The unique geographical location of Hainan means the climate characteristics are different from those of inland areas in China.[11] As China's only tropical province, most of the island has a humid tropical monsoon climate characterised by hot and humid summers with mild, pleasant winters.[12] The annual average temperature of Hainan usually ranges from 24°C to 35°C in the summer and ftom 19°C to 25°C in the winter.[13] Such a climate could contribute to the proliferation and spread of micro-organisms, such as Epstein-Barr virus.[14] Most Hainan residents consume a light diet, with fewer oils, a sweeter taste and less seasoning than diets on the Chinese mainland.[15] Seafood, coffee and tea are the most common items in the Hainanese diet.[16] A

previous study reported that the prevalence of NCDs in Haikou (capital city of Hainan) was significantly different from those in other parts of mainland China.[17] This may be caused by different environmental exposures or lifestyles between Hainan residents and the mainland Chinese population. Therefore, analysing the influence of Hainan's climate, dietary habits and lifestyle on NCDs is of important guiding significance for the health of Hainan residents. In addition, inherited factors may play a critical role in the progression of NCDs in Hainan residents. Although genome-wide association studies have led to successful identification of a large number of common gene variants associated with NCDs,[18–22] these findings may not be appropriate for the Hainan population, whose physical characteristics are different from those of most of the mainland Chinese population. Furthermore, studies that provide clear explanations of the effects of potential interactions between genetic and unique environmental risk factors on NCDs are scarce.[23 24] To date, there is no independent biospecimen-based prospective study on NCDs in Hainan. Aetiological descriptions and prevention and control measures for NCDs in mainland Chinese and Western populations cannot be generalised to the Hainan population. There is an urgent need to establish a large, localised population-based cohort; this not only has important practical significance for the health of people in Hainan but also will provide a scientific reference for neighbouring island countries and regions, such as Southeast Asia.

From June 2018 to November 2020, we conducted a large population-based prospective cohort study in Hainan, with an appropriate study design and rigorous data quality control. The main study objectives of the Hainan Cohort were to (1) establish a large population-based cohort database and biobank; (2) explore the prevalence and incidence of and risk factors for common NCDs; (3) clarify the relationships among environmental factors, lifestyle, psychosocial factors and NCDs, and potential effect modifications among these factors; (4) assess the effects of genetic factors and gene–environment interactions on NCDs; (5) identify potential biomarkers of NCDs using metabolomic analysis; and (6) establish a tracking system to monitor the occurrence, progression and prognosis of common NCDs, and all-cause death; and[7] develop and validate a health risk assessment and risk prediction model for NCDs.

COHORT DESCRIPTION
Selection of study site and participant eligibility
The Hainan Cohort study was designed as a longitudinal study that recruited community dwellers in five main areas of Hainan (figure 1): Haikou (the capital city located on the northern coast of Hainan), Sanya (central city in the south of Hainan), Qionghai (a city located in the east of Hainan), Dongfang (a city located on the western coast of Hainan) and Qiongzhong (Li and Miao Autonomous County in central Hainan). These areas cover both urban
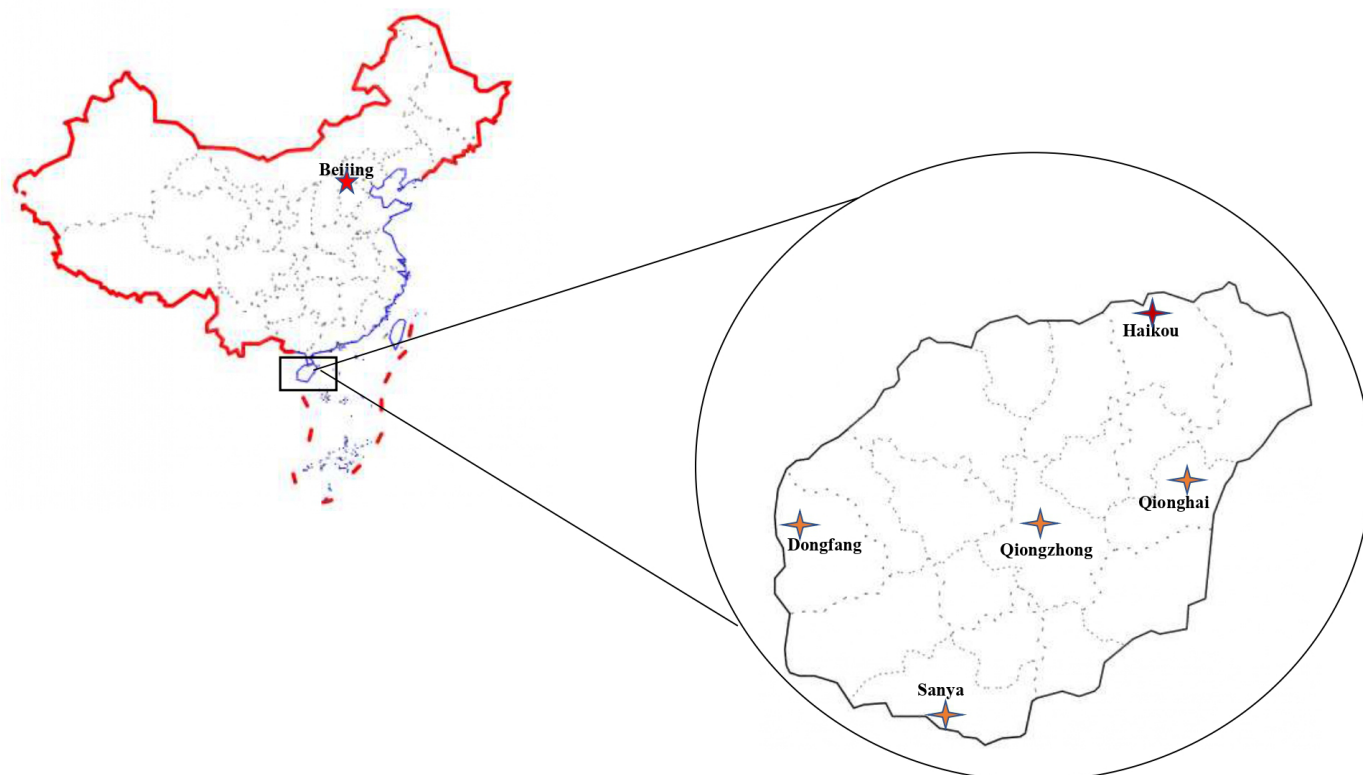

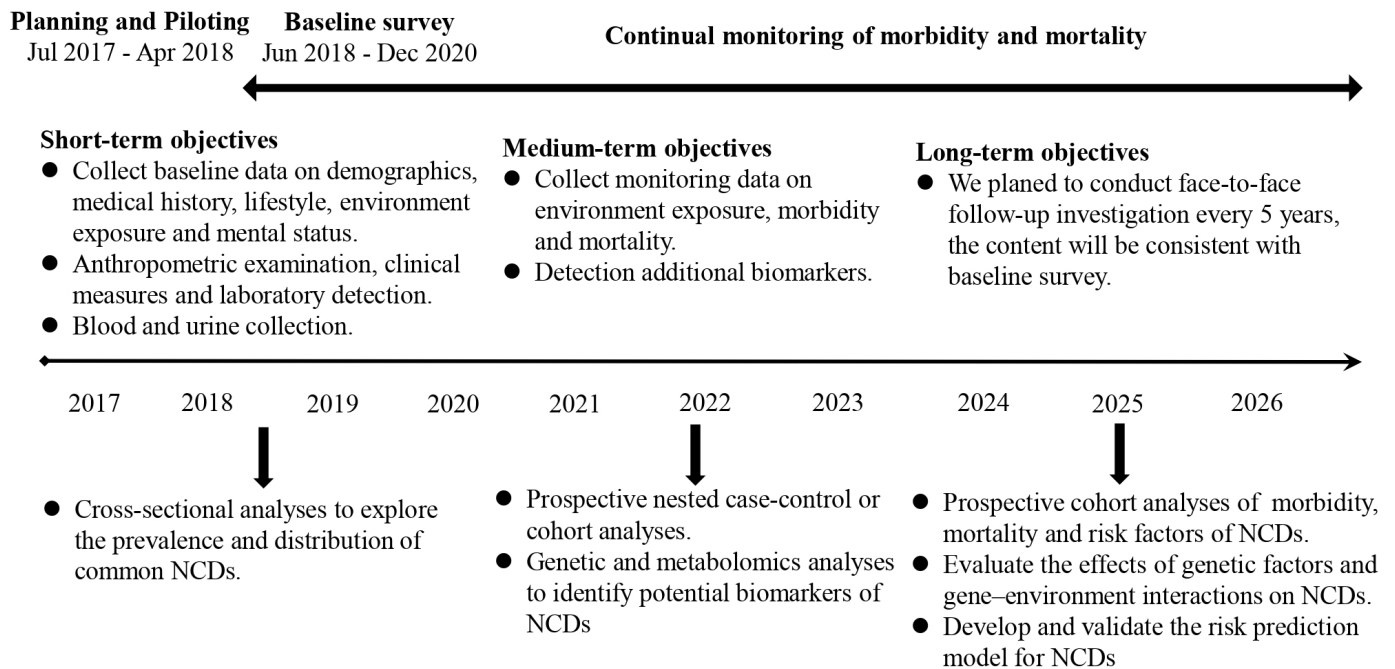**Figure 1** Locations of the five survey sites in the Hainan Cohort.

Figure 2 Study plan of the Hainan Cohort study. NCD, non-communicable disease.

and rural locations, as well as different ethnic groups who speak different dialects, and are representative of the diverse sociodemographic characteristics of Hainan. We determined the sample size of each area based on the population distribution of Hainan and other objective factors, such as biological sample preservation, transportation and local coordination. Then, a multistage cluster random-sampling method was used to obtain samples from the general population. In the first stage, 18 districts were selected from five major areas through simple cluster sampling. Second, communities or units in each district were randomly selected considering population stability, local medical conditions and adherence to future follow-up assessments. The inclusion criteria were as follows: (1) age between 35 years and 74 years old; (2) residence in Hainan for at least 5 years; (3) no severe physical disability, mental illness or severe medical condition; and (4) ability to complete questionnaire interviews, physical examinations and follow-up assessments. All the participants enrolled in this study voluntarily and signed

Table 1	Summary of baseline measures in the Hainan Cohort study
Exposure category	**Variable/exposure**
General demographic information	Name, ID card number, home address Sex, birthday, marital status, occupation, years of working and annual household income
Medical history	Diagnosed diseases (diabetes mellitus, coronary heart disease, stroke, hypertension, hyperuricaemia, cancer, chronic bronchitis, pulmonary tuberculosis, asthma, chronic hepatitis and other digestive system disease), age at diagnosis, current treatment, use of medicines during the last 6 months and the site of cancer
Family history	Family (parents, siblings and children) history of coronary heart disease, hypertension, stroke, diabetes mellitus, cancer and chronic obstructive pulmonary disease
Lifestyle habits	Smoking status, alcohol consumption, tea consumption, dietary pattern (frequency and type of breakfast and midnight snacks, frequency of major food* consumption, frequency and type of soup consumption and supplement use), passive smoking and indoor pollution, physical activities (IPAQ), defecation status, sleeping status (PSQI) and napping
Female reproductive history	Menstruation status (ages at first menstruation and menopause), history of pregnancy, history of breast feeding, history of contraception, history of gynaecological diseases and history of gynaecological surgery
Mental status	CIDI-A, CIDI-B

*Major foods includes rice, wheat, corn, dessert, meat, poultry, sea or fish food, fresh eggs, fresh vegetables, fresh fruit, nuts, marinated vegetables or meat, dairy products and beverages.
CIDI, Comprehensive International Diagnostic Interview; IPAQ, International Physical Activity Questionnaire; PSQI, Pittsburgh Sleep Quality Index.

Table 2 Summary of anthropometric and clinical measurements collected at baseline in the Hainan Cohort study

Variables	Measurements (n)	Equipment used
Standing height	One	SK-X80; Shuangjia, China
Body weight	One	BC-6017611; Tanita, Japan
Body fat percentage	One	BC-6017611; Tanita, Japan
Basal metabolic rate	One	BC-6017611; Tanita, Japan
Visceral fat index	One	BC-6017611; Tanita, Japan
Waist circumference	One	Hoechstmass, China
Hip circumference	One	Hoechstmass, China
Resting blood pressure	Two	HEM-7071; OMRON, Japan
Heart rate	Two	HEM-7071; OMRON, Japan
Vital capacity	Two	SP10; Contec, China
Lung carbon monoxide detector	Two	PICO; Philips, USA
12-lead electrocardiography	One	ECG-1350C; NEC Lighting, China
Fasting plasma glucose	One	7600–110; Hitachi, Japan
Blood lipids	One	7600–110; Hitachi, Japan
Hepatic function	One	7600–110; Hitachi, Japan
Kidney function	One	7600–110; Hitachi, Japan
Blood routine (24 items)	One	BC-6900; Mindray, China
Urine routine	One	COBIO SCAN XL; Kobold, Germany
Urinary albumin	One	COBIO SCAN XL; Kobold, Germany
Urine creatinine	One	COBIO SCAN XL; Kobold, Germany

an informed consent form on enrolment. A unique identification was created for each participant using their national identification card.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Study plan

The Hainan Cohort is a dynamic cohort with no end date, and we will collect follow-up data for each participant periodically until death. Follow-up will be conducted using two methods: monitoring and long-term follow-up surveys. For basic monitoring of the participants, we are cooperating with the Social Insurance Service Centre and Healthcare Security Administration of Hainan. Using the independent medical insurance number of each urban and rural participant who signed the informed consent form, information about medical treatment, we will collect hospitalisation and incidence of diseases from the department of medical insurance data management in January and July of each year. The Mortality Surveillance System of the Hainan Centre for Disease Control and Prevention allows us to track cause-specific mortality in all the participants in the cohort. Causes of death will be classified according to the 10th version of the International Statistical Classification of Diseases and Related Health Problems. The National Basic Public Health Service Platform (http://www.nbphsp.org.cn/) has achieved full coverage in Hainan, enabling us to ascertain the onset time of common chronic diseases, such as hypertension and diabetes, in the cohort.

Long-term follow-up will be conducted every 5 years. All surviving participants will be visited and will complete an in-person assessment, which is the same as that conducted at baseline, that includes a questionnaire interview, physical examination, and blood and urine sample collection. The onset of endpoint events and the loss of cohort participants will be investigated and compared with the monitoring data to reduce possible omissions and misstatements caused by conventional monitoring. Endpoint events for major chronic diseases will be confirmed by experienced physicians based on well-accepted international standards. The household registration system of the public security department has the relocation records of all permanent residents, which can be combined with the national network medical insurance system to follow up any study participants who can still be contacted after relocation and finally control the rate of lost to less than an estimated 8%. Figure 2 shows the overall study plan.

Baseline assessments

In December 2017, a pilot study was carried out before the formal investigation. The aim of the pilot study was to evaluate whether each item on the questionnaire was easy to understand by persons in different age groups and from different cultural backgrounds and whether the options were logical. Additionally, the pilot study allowed

Table 3 Baseline characteristics of the participants in the Hainan Cohort study

Variables	Total (n=14 443)	Men (n=7484)	Women (n=6959)
Age (years)	48.8 (9.6)	49.5 (9.3)	48.1 (9.8)
Age group			
35~39	2647 (18.3)	1301 (17.4)	1346 (19.3)
40~44	2752 (19.1)	1219 (16.3)	1533 (22.0)
45~49	3147 (21.8)	1372 (18.3)	1775 (25.5)
50~54	2273 (15.7)	1403 (18.7)	870 (12.5)
55~59	1612 (11.2)	1244 (16.6)	368 (5.3)
60~64	646 (4.5)	339 (4.5)	307 (4.4)
65~69	780 (5.4)	331 (4.4)	449 (6.5)
70~74	586 (4.1)	275 (3.7)	311 (4.5)
Education level			
Illiterate	523 (5.2)	146 (2.9)	377 (7.5)
Primary school	1120 (11.1)	406 (7.9)	714 (14.2)
Middle school	2946 (29.1)	1359 (26.6)	1587 (31.6)
High school	2788 (27.5)	1670 (32.7)	1118 (22.3)
University or college or higher	2757 (27.2)	1532 (30.0)	1225 (24.4)
Marital status			
Married	9001 (89.7)	4585 (90.6)	4416 (88.7)
Widowed	332 (3.3)	122 (2.4)	210 (4.2)
Divorced	348 (3.5)	155 (3.1)	193 (3.9)
Single	358 (3.6)	200 (4.0)	158 (3.2)
Annual household income (¥)			
<10 000	325 (3.3)	144 (2.9)	181 (3.7)
10 000–29 999	2131 (21.7)	901 (18.3)	1230 (25.2)
30 000–59 999	3263 (33.3)	1695 (34.4)	1568 (32.1)
60 000–99 999	2187 (22.3)	1208 (24.5)	979 (20.0)
>100 000	1899 (19.4)	974 (19.8)	925 (18.9)
Smoking			
Never	7392 (72.4)	2394 (46.4)	4998 (98.9)
Current	2416 (23.7)	2364 (45.8)	52 (1.0)
Former	405 (4.0)	401 (7.8)	4 (0.1)
Alcohol consumption			
Never	7221 (77.1)	3185 (64.9)	4036 (90.5)
Current	1794 (19.2)	1507 (30.7)	287 (6.4)
Former	352 (3.8)	217 (4.4)	135 (3.0)
Passive smoking	1856 (24.5)	422 (16.0)	1434 (29.1)
Soup consumption			
Everyday	1477 (14.9)	828 (16.8)	649 (13.1)
Sometimes	4089 (41.4)	1929 (39.0)	2160 (43.7)
Never	4320 (43.7)	2184 (44.2)	2136 (43.2)
Skips breakfast			
Everyday	541 (5.3)	269 (5.2)	272 (5.4)
Sometimes	1906 (18.8)	952 (18.6)	954 (19.0)
Never	7715 (75.9)	3909 (76.2)	3806 (75.6)
Consumes a midnight snack			
Everyday	376 (3.7)	188 (3.7)	188 (3.7)
Sometimes	1694 (16.7)	910 (17.7)	784 (15.6)

Continued

Table 3 Continued

Variables	Total (n=14443)	Men (n=7484)	Women (n=6959)
Never	8090 (79.6)	4035 (78.6)	4055 (80.7)
Regular exercise	2676 (27.9)	1400 (29.1)	1276 (26.6)
Sleep duration (hour)	7.2 (1.3)	7.1 (1.3)	7.2 (1.3)
BMI (kg/m^2)	23.7 (3.5)	24.1 (3.5)	23.3 (3.5)
Waist circumference (cm)	82.1 (9.7)	85.6 (9.5)	78.7 (8.6)
Hip circumference (cm)	94.1 (7.9)	95.1 (8.0)	93.2 (7.6)
Systolic blood pressure (mm Hg)	130.2 (20.0)	133.7 (19.6)	126.6 (19.8)
Diastolic blood pressure (mm Hg)	82.1 (13.5)	85.6 (13.5)	78.5 (12.5)
Heart rate (beats/min)	79.7 (11.9)	79.5 (12.3)	80.0 (11.4)
Body fat rate (%)	26.8 (14.1)	21.7 (10.9)	31.9 (15.0)
Skeletal muscle (kg)	43.8 (55.1)	49.8 (16.8)	37.7 (75.4)
Bone mass (kg)	2.7 (5.5)	3.1 (7.4)	2.3 (2.0)
Perspiration rate (%)	54.9 (108.2)	58.1 (123.7)	51.8 (90.1)
Visceral fat index	8.4 (4.9)	10.7 (5.3)	6.0 (3.0)

*Continuous variables are presented as means and SD, and categorical variables are presented as counts and percentages.
BMI, body mass index.

us to find the most suitable method for asking difficult questions after discussion to facilitate the training of investigators in the future. The baseline assessments were performed from June 2018 to October 2020 and included a questionnaire survey, physical examination, blood and urine sample collection, laboratory measurements and outdoor environmental data collection.

After overnight fasting, all participants were required to bring their ID cards or household registration books to the local health examination centre at the specified time. Following the obtainment of written informed consent, a validated face-to-face questionnaire interview was conducted by well-trained interviewers from the School of Public Health, Hainan Medical University. The questionnaire collected general demographic (name, sex, age, ethnicity, education level, marital status, occupation and annual household income), medical history (diabetes mellitus, coronary heart disease, stroke, hypertension, hyperuricaemia, cancer, etc), family disease history (coronary heart disease, hypertension, stroke, diabetes mellitus, cancer and chronic obstructive pulmonary disease), lifestyle habit (smoking, alcohol consumption, tea consumption, dietary pattern, sleeping status, physical activity, passive smoking and indoor pollution), female reproductive history and mental status data. Table 1 shows detailed information of the baseline measures.

Physical examinations were conducted by trained doctors or nurses from the First Affiliated Hospital of Hainan Medical University; collected measurements included height, waist circumference, hip circumference, blood pressure, heart rate, lung function, body composition (weight, body mass index (BMI), body fat percentage, skeletal muscle mass, bone mass, daily energy requirement, etc), as well as 12-lead electrocardiography. All measurements were collected by trained investigators

in a large, bright room with a suitable temperature using standard procedures. Height was measured without shoes and hats using an ultrasonic height measuring instrument (SK-X80, China). Waist and hip circumferences were measured while the participants were wearing thin clothing. The investigator stood on the right side of the participant and measured them with a tape measure (Hoechstmass, China). The above three measurements were accurate to the nearest 0.1 cm. Blood pressure and heart rate were measured twice consecutively using a validated electronic sphygmomanometer (HEM-7071; OMRON, Japan) on the right arm after the participants had rested for at least 5 min in a sitting position. Forced expiratory volume in 1 s and forced vital capacity were measured using a spirometer (SP10; Contec, China); the participants took a deep breath and blew at the fastest speed and with maximum strength until all the gas in their lungs was exhaled. Body composition was measured using a body fat detector (BC-6017611; Tanita, Japan). Participants removed their coats, shoes and socks and checked that the soles of their feet were clean; when necessary, they used a wet towel to wipe their feet, and they removed all metal objects (such as mobile phones and keys) to avoid affecting the result. Weight was measured to the nearest 0.1 kg. Electrocardiography was performed (ECG-1350C; NEC Lighting, China), and the experienced physician reviewed each ECG. BMI was calculated as weight (kg) divided by height (m) squared. Table 2 shows the details of the equipment for the physical examination and clinical measurements.

Before blood collection, all participants fasted for at least 10 hours as required. To prevent the occurrence of hypoglycaemic events, all blood samples were collected between 07:00 and 10:00. Three tubes of peripheral venous blood from each participant were

Table 4 Baseline levels of biochemical traits of the participants in the Hainan Cohort study

Variables	Total	Men	Women
Lipids			
TGs (mmol/L)	1.6 (1.8)	2.0 (2.4)	1.2 (0.8)
TC (mmol/L)	5.5 (3.8)	5.6 (5.2)	5.4 (1.2)
HDL-C (mmol/L)	1.9 (1.0)	2.0 (1.2)	1.9 (0.8)
LDL-C (mmol/L)	3.2 (2.2)	3.3 (1.5)	3.1 (2.8)
Fasting glucose (mmol/L)	5.7 (1.7)	5.8 (1.9)	5.5 (1.3)
Renal function			
Creatinine (µmol/L)	69.4 (21.9)	80.1 (23.3)	57.6 (12.2)
UA (µmol/L)	325.8 (91.2)	373.9 (83.4)	270.7 (64.7)
Hepatic function			
ALT (µ/L)	5.3 (14.8)	5.6 (13.5)	4.9 (16.1)
AST (µ/L)	26.6 (24.5)	28.8 (19.6)	24.3 (28.8)
Total protein (g/L)	74.6 (3.7)	74.3 (3.8)	74.9 (3.5)
ALB (g/L)	43.1 (4.1)	43.7 (3.9)	42.3 (4.1)
ALP (µ/L)	71.7 (20.7)	75.9 (20.6)	68.5 (20.3)
GGT (µ/L)	47.0 (83.0)	67.2 (106.2)	23.8 (29.6)
Globulin (g/L)	30.8 (3.5)	29.9 (3.6)	31.8 (3.1)
Total bilirubin (µmol/L)	2.8 (1.5)	3.2 (1.7)	2.4 (1.1)
Direct bilirubin (µmol/L)	8.5 (4.1)	9.4 (4.4)	7.5 (3.6)
Routine blood			
WBC count (×10^9/L)	6.7 (1.9)	7.0 (1.8)	6.3 (1.9)
RBC count (×10^9/L)	5.1 (0.7)	5.3 (0.7)	4.8 (0.6)
Haemoglobin (g/L)	143.5 (20.1)	154.5 (15.4)	131.2 (17.4)
PLT count (g/L)	269.5 (75.4)	255.8 (67.8)	284.8 (80.5)
RCDW	13.3 (4.0)	13.3 (4.6)	13.2 (3.2)

*Continuous variables are presented as means and SD, and categorical variables are presented as counts and percentages.
ALB, albumin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; FPG, fasting plasma glucose; GGT, gamma-glutamyl transpeptidase; HbA1c, glycated haemoglobin; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; PLT, platelet; RBC, red blood cell; RCDW, red cell distribution width; TC, total cholesterol; TG, triglyceride; WBC, white blood cell.

extracted for a total of 15 mL; two tubes were EDTA anticoagulation tubes for the separation of plasma and blood cells, and one tube was a non-anticoagulation tube for the separation of serum and clotted blood. The collected blood samples were sent to the laboratory for testing within 2 hours. The serum samples were used to detect biochemical indicators with automatic biochemical analyser (7600–110; Hitachi, Japan), including alanine aminotransferase, aspartate aminotransferase, albumin, alkaline phosphatase, gamma-glutamyl transpeptidase, triglycerides (TGs), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), urea, creatinine, uric acid (UA), fasting plasma glucose (FPG) and glycated haemoglobin. Routine blood parameters, including white blood cell count, red blood cell count, red cell distribution width, platelet count, etc, were measured using

an automatic blood cell analyser (BC-6900; Mindray, China). The remaining blood samples were stored in a 4°C refrigerator and centrifuged within 1 day after collection. Plasma, serum, and white and red cells were collected after 15 min of centrifugation at 3000 rpm. Aliquots of 0.2 mL of centrifuged blood samples were retained in each tube and stored in a −80°C refrigerator to prevent DNA breakage or degradation. At least 50 mL of midstream morning urine on an empty stomach was collected using a urine cup and immediately placed into a refrigerator at −4°C after collection. After testing, the remaining urine samples were divided into two 15 mL centrifuge tubes using disposable plastic straws (10–15 mL each) and sent back to the laboratory for preservation within 24 hours. The routine urine test indicators included urine glucose, urine protein, urobilirubin, etc. All these test items were measured by the general standardised methods
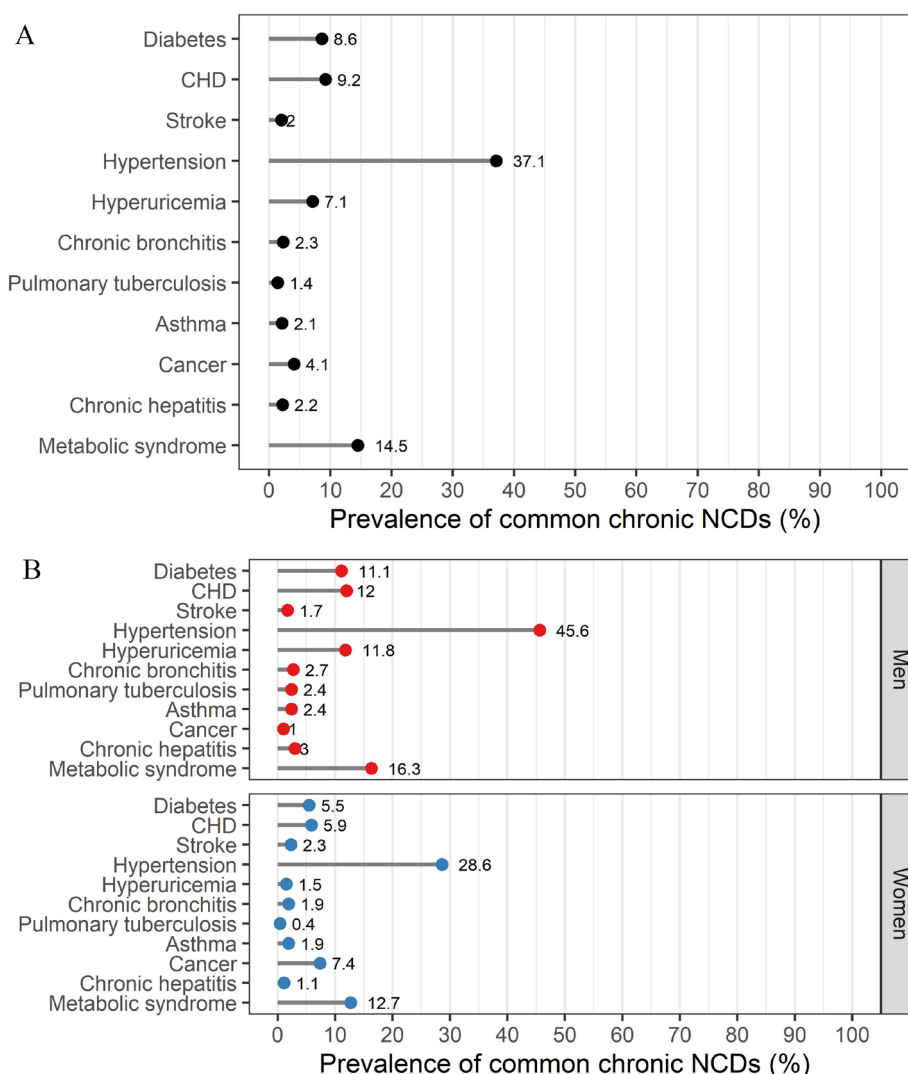
Figure 3 Prevalence of common chronic NCDS at baseline in the Hainan Cohort study. (A) Total participants, (B) stratified by sex. CHD, coronary heart disease; NCD, non-communicable disease.

of the Clinical Laboratory of the Affiliated Hospital of Hainan Medical University.

Outdoor environmental data, including daily maximum temperature, daily minimum temperature and average temperature, were collected from the National Meteorological Bureau and local meteorological departments. Air quality monitoring data at the survey site, including the daily air quality index, air quality levels and primary pollutant (SO_2, NO_2, CO, O_3, PM_{10} and $PM_{2.5}$) concentrations, were obtained through the air quality platform of the National and Hainan Environmental Protection Departments.

Hypertension was defined as a sitting systolic blood pressure (average of 2 readings with a difference of <5 mm Hg) of 140 mm Hg or diastolic blood pressure of 90 mm Hg and/or the use of antihypertensive drugs.[25] Diabetes mellitus was defined as a fasting blood glucose value of ≥7.0 mmol/L, prior diagnosed diabetes mellitus or the use of antidiabetic agents.[26] Metabolic syndrome (MS) was defined according to the criteria of the International Diabetes Federation (IDF) for Chinese adults. MS was defined according to the IDF criteria and the Chinese specific abdominal obesity standard in adults (waist circumference in Chinese men ≥90 cm, waist circumference in Chinese women ≥80 cm), along with the presence of two or more of the following: (1) TG level of >150 mg/dL (1.7 mmol/L) or drug treatment for elevated TGs, (2) HDL level of <40 mg/dL (0.9 mmol/L) in men, < 50 mg/dL (1.1 mmol/L) in women or drug treatment for low HDL; (3) systolic blood pressure of ≥130 mm Hg or diastolic blood pressure of ≥85 mm Hg or drug treatment for hypertension, and (4) FPG of ≥100 mg/dL (5.6 mmol/L) during fasting status or diagnosed diabetes.[27]

In addition, a resurvey was conducted in a random subsample of the Hainan Cohort during the time of the baseline survey. Approximately 5% of participants were randomly chosen and completed the resurvey. The procedures and content of the resurvey were consistent with those of the baseline survey.

Table 5 Spearman correlation coefficients for selected measurements between the baseline survey and resurvey among 637 participants

Variable	Total	Men	Women
Physical measurements			
BMI	0.81	0.96	0.67
Waist circumference	0.83	0.86	0.74
Hip circumference	0.57	0.51	0.69
Systolic blood pressure	0.75	0.78	0.68
Diastolic blood pressure	0.69	0.75	0.56
Biochemical test			
TGs	0.61	0.42	0.96
TC	0.79	0.77	0.81
Fasting glucose	0.84	0.84	0.81
Creatinine	0.90	0.83	0.80
UA	0.89	0.84	0.82
Blood routine			
WBC count	0.94	0.92	0.96
RBC count	0.97	0.98	0.92
Haemoglobin	0.98	0.97	0.96
Platelet count	0.98	0.98	0.98

BMI, body mass index; RBC, red blood cell; TC, total cholesterol; TGs, triglyceride; UA, uric acid; WBC, white blood cell.

FINDING TO DATE

Of the 16070 eligible people, 14602 participants were enrolled and underwent at least one major measurement at baseline, with a participation rate of 90.1%. The most common reasons for declining participation in the study were no interest in or lack of understanding of the study, no time to participate in the examinations and fear of exsanguination. Of those who completed the survey, 159 were excluded because they were younger than 35 or older than 74 years at baseline. Finally, a total of 14443 participants fulfilled the inclusion criteria of this study. The baseline characteristics of the participants are presented in table 3. Among the participants, 51.8% (n=7484) were men; the mean age was 48.8 (SD 9.6) years; and less than 15% of people were over the age of 60. A total of 83.7% of the participants had a secondary school education or above, and the illiteracy rate was only 5.2%. The proportion of married participants was 89.7% for both sexes. Divorced, widowed and single participants accounted for 3.3%, 3.5% and 3.6%, respectively. The most commonly reported annual household income level was between ¥30 000 and ¥60 000 (33.3%). Among men, 45.8% were current smokers, and 7.8% were former smokers. Among women, 98.9% of participants had never smoked. Nevertheless, the percentage of women suffering from passive smoke was higher than that of men (29.1 vs 16.0%). Regarding alcohol consumption, 30.7% of men and 6.4% of women were current alcohol consumers, and 64.9% of men and 90.5% of women were never alcohol consumers. More than half of the participants consumed soup (41.4% consumed it sometimes, 14.9% consumed it every day). A total of 5.3% of the participants never ate breakfast, and 3.7% of participants ate midnight snacks every day. Regarding regular exercise, the prevalence in men was slightly higher than that in women (29.1% of men and 26.6% of women reported engaging in regular exercise). The mean sleep duration, BMI, waist and hip circumferences, systolic blood pressure and diastolic blood pressure, heart rate, body fat rate, skeletal muscle and bone mass, body moisture rate and visceral fat index were 7.2 hours, $23.7 \, kg/m^2$, 82.1 and 94.1 cm, 130.2 and 82.1 mm Hg, 79.7 beats/min, 26.8%, 43.8 and 2.7 kg, 54.9% and 8.4, respectively.

Table 4 shows the biochemical characteristics of the participants at baseline. The mean TG, TC, HDL-C, LDL-C, FPG, creatinine and UA levels were 1.6 mmol/L, 5.5 mmol/L, 1.9 mmol/L, 5.7 mmol/L, 69.4 μmol/L and 325.8 μmol/L, respectively. Information about other hepatic function and routine blood results can be found in table 4.

The baseline prevalence of common chronic NCDs is presented in figure 3. The prevalence rates of diabetes, coronary heart disease, stroke, hypertension, hyperuricaemia, chronic bronchitis, pulmonary tuberculosis, asthma, cancer, chronic hepatitis and MS were 8.6%, 9.2%, 2.0%, 37.1%, 7.1%, 2.3%, 1.4%, 2.1%, 4.1%, 2.2% and 14.5%, respectively. Men had higher rates of diabetes, coronary heart disease, hypertension, hyperuricaemia, chronic bronchitis, pulmonary tuberculosis, asthma, chronic hepatitis and MS than women, while women had higher rates of stroke and cancer.

A total of 637 (4.4%) participants completed the resurvey. Table 5 shows the correlation of some of the variables between the baseline survey and the resurvey. For most of the selected measurements, the resurvey was in good concordance with the baseline survey. Among these measurements, the correlation coefficient of routine blood indexes was the highest, at approximately 0.95. The correlation coefficient between physical examination, and blood biochemical examination results ranged between approximately 0.6 and 0.9.

Strengths and limitations

This cohort study has several clear strengths. First, as far as we know, the Hainan Cohort study is the first independent biospecimen-based prospective study on NCDs in tropical China to analyse detailed information including sociodemographic characteristics, lifestyle habits, personal and family histories of diseases, dietary pattern, sleeping status, physical activity levels, female reproductive history, and anthropometric and clinical measurements. Second, the sample size was relatively large, which gives us sufficient power to explore the relationship between risk factors and common chronic NCDs, such as diabetes mellitus, chronic obstructive pulmonary disease, myocardial infarction and stroke (see sample size estimation in online supplemental materials 1). Third, the

sample biospecimens collected at baseline will allow us to conduct nested case–control studies on genetic variations and metabolomic biomarkers for NCDs. Fourth, due to rigorous data quality control, the age distribution and sex ratio of the baseline data can generally be representative of those in the Chinese population (based on the seventh Chinese National Census). Finally, we linked the cohort database with the medical insurance system, the mortality surveillance system and the national basic public health service platform and will combine these data with those obtained from long-term follow-up surveys to ensure that we obtain accurate endpoint information.

This study also has limitations. First, we did not collect the basic data of eligible people who did not participate in the survey according to the sampling plan, which may lead to selection bias. Second, for some NCDs with a low incidence rate (such as lung cancer), the sample size of the study is insufficient. We will consider expanding the number of participants in the next wave of follow-up. Third, although we carefully checked the collected questionnaires and the resurvey showed good concordance with the baseline survey, we could not rule out possible recall and measurement bias during data collection. Finally, more than 98% of the participants in the cohort were Han Chinese; thus, caution should be taken when generalising the results of this work to other ethnicities.

ORCID iD
Fan Zhang http://orcid.org/0000-0001-9600-2737

REFERENCES

1 Martinez R, Lloyd-Sherlock P, Soliz P, *et al*. Trends in premature avertable mortality from non-communicable diseases for 195 countries and territories, 1990-2017: a population-based study. *Lancet Glob Health* 2020;8:e511–23.

2 Heller O, Somerville C, Suggs LS, *et al*. The process of prioritization of non-communicable diseases in the global health policy arena. *Health Policy Plan* 2019;34:370–83.

3 Savage A, McIver L, Schubert L. Review: the nexus of climate change, food and nutrition security and diet-related non-communicable diseases in Pacific Island countries and territories. *Clim Dev* 2020;12:120–33.

4 Tohidinezhad F, Khorsand A, Zakavi SR, *et al*. The burden and predisposing factors of non-communicable diseases in Mashhad University of medical sciences personnel: a prospective 15-year organizational cohort study protocol and baseline assessment. *BMC Public Health* 2020;20:1637.

5 Budreviciute A, Damiati S, Sabir DK, *et al*. Management and prevention strategies for non-communicable diseases (NCDS) and their risk factors. *Front Public Health* 2020;8:574111.

6 Pan T, Palmer M. Risk factors and non-communicable disease diagnosis in China. *China Economic Review* 2018;50:72–84.

7 Zhou M, Wang H, Zeng X, *et al*. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2019;394:1145–58.

8 Liu S, Li Y, Zeng X, *et al*. Burden of cardiovascular diseases in China, 1990-2016: findings from the 2016 global burden of disease study. *JAMA Cardiol* 2019;4:342–52.

9 Xu Y, Wang L, He J, *et al*. Prevalence and control of diabetes in Chinese adults. *JAMA* 2013;310:948–59.

10 Xiong W, Wang Q, Xie D, *et al*. Factors influencing tropical Island freshwater fishes: species, status, threats and conservation in Hainan Island. *Knowl Manag Aquat Ecosyst* 2018;16:6.

11 Lu S, Pang B, Qi Y, *et al*. Field study of thermal comfort in non-air-conditioned buildings in a tropical island climate. *Appl Ergon* 2018;66:89–97.

12 Wang L, Zhang L, Lv J, *et al*. Public awareness of drinking water safety and contamination accidents: a case study in Hainan Province, China. *Water* 2018;10:446.

13 Leou EC, JX L. Target Duo: the project of "international tourism island" in hainan special economic zone, China. In: Li Y, Leandro F, Figueiredo PG, eds. *Handbook of research on special economic zones as regional development enablers. Hershey*. Pennsylvania: IGI Global, 2022: 206–33.

14 Inoue Y, Yazawa A, Li D, *et al*. Epstein-Barr virus antibody titer and its association with the domain scores from the World Health Organization's quality of life questionnaire: findings from rural Hainan Province, China. *Am J Hum Biol* 2014;26:51–5.

15 Peng Z, Wang Y, Huang X, *et al*. Dietary vitamin intake and risk of metabolic syndrome among centenarians in China. *Exp Ther Med* 2021;21:105.

16 Yang S, Wang S, Wang L, *et al*. Dietary behaviors and patterns of centenarians in Hainan: a cross-sectional study. *Nutrition* 2021;89:111228.

17 Zhao C, Wong L, Zhu Q, *et al*. Prevalence and correlates of chronic diseases in an elderly population: a community-based survey in Haikou. *PLoS One* 2018;13:e0199006.

18 Chiang K-M, Chang H-C, Yang H-C, *et al*. Genome-wide association study of morbid obesity in Han Chinese. *BMC Genet* 2019;20:97.

19 Liu X, Wang Q, Zhu R. Association of GWAS-susceptibility loci with ischemic stroke recurrence in a Han Chinese population. *J Gene Med* 2021;23:e3264.

20 Hu Z, Shi Y, Mo X, *et al*. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat Genet* 2013;45:818–21.

21 Zhang H, Zhai Q, Zhang Z, *et al*. Association of GWAS-supported variants rs556621 on chromosome 6p21.1 with large artery atherosclerotic stroke in a southern Chinese Han population. *Neuromolecular Med* 2017;19:94–100.

22 Hu Z, Wu C, Shi Y, *et al*. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet* 2011;43:792–6.

23 Moore JH, Asselbergs FW, Williams SM. Bioinformatics challenges for genome-wide association studies. *Bioinformatics* 2010;26:445–55.

24 Yap RWK, Shidoji Y, Yap WS, *et al*. Association and interaction effect of AGTR1 and AGTR2 gene polymorphisms with dietary pattern on metabolic risk factors of cardiovascular disease in Malaysian adults. *Nutrients* 2017;9:853.

25 Wang Z, Chen Z, Zhang L, *et al*. Status of hypertension in China: results from the China hypertension survey, 2012-2015. *Circulation* 2018;137:2344–56.

26 Li Y, Teng D, Shi X, *et al*. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American diabetes association: national cross sectional study. *BMJ* 2020;369:m997.

27 Alberti KGMM, Eckel RH, Grundy SM, *et al*. Harmonizing the metabolic syndrome: a joint interim statement of the International diabetes Federation task force on epidemiology and prevention; National heart, lung, and blood Institute; American heart association; world heart Federation; international atherosclerosis society; and international association for the study of obesity. *Circulation* 2009;120:1640–5.