



Cohort Profile

Cohort Profile: the China Multi-Ethnic Cohort (CMEC) study

Xing Zhao,^{1*} Feng Hong,² Jianzhong Yin,^{3,4} Wenge Tang,⁵ Gang Zhang,⁶ Xian Liang,⁷ Jingzhong Li,⁸ Chaoying Cui,⁹ and Xiaosong Li¹ on behalf of the China Multi-Ethnic Cohort (CMEC) collaborative group

¹West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China, ²School of Public Health, the key Laboratory of Environmental Pollution Monitoring and Disease Control, Ministry of Education, Guizhou Medical University, Guiyang 550025, China,, ³Kunming Medical University, Kunming, Yunnan, China, ⁴Baoshan College of Traditional Chinese Medicine, Baoshan, Yunnan, China, ⁵Chongqing Centre for Disease Control and Prevention, Chongqing, China, ⁶Sichuan Provincial Hospital for Women and Children (Affiliated Women and Children's Hospital of Chengdu Medical College), Chengdu, China, ⁷Chengdu Centre for Disease Control and Prevention, Chengdu, Sichuan, China, ⁸Tibet Centre for Disease Control and Prevention, Lhasa, Tibet, China and ⁹High Altitude Medical Research Centre, School of Medicine, Tibet University, Lhasa, Tibet, China

*Corresponding author. West China School of Public Health and West China Fourth Hospital, Sichuan University, No. 16, Section 3, Renmin South Road, Chengdu, Sichuan, CN 610041, China. E-mail: xingzhao@scu.edu.cn

Editorial decision 18 August 2020; Accepted 27 August 2020

Why was the cohort set up?

Non-communicable diseases (NCDs), such as cardiovascular disease, cancer, chronic respiratory disease and diabetes, have become leading causes of death worldwide.^{1,2} It is estimated that 63% of all global deaths are due to NCDs,³ with an increasing contribution from low- and middle-income countries. The prevalence and incidence of NCDs are not evenly distributed between and within countries because of variations in environment, lifestyle, socio-economic position, genetic factors and access to health care services.^{4,5}

China has been facing a growing burden of NCDs with strong regional variation.⁶ Most primary data on the prevalence of and risk factors for NCDs come from Central and Eastern China,^{7,8} with a mix of ethnicity, socio-economic status, geography, and dietary, behavioural and lifestyle factors that is very different compared to Western

China. Ethnicity in particular is increasingly seen as an important risk factor for NCDs throughout the world.⁹ Ethnic minority groups are known to have different outcomes for cancer,¹⁰ diabetes,¹¹ cardiovascular diseases⁹ and respiratory diseases¹² compared with ethnic majority groups.

China, with its 55 different ethnic minority groups, represents the largest ethnic minority population in the world (114 million in 2010).¹³ However, remarkably little is known about ethnic variation in the prevalence, incidence and/or risk factors for NCDs in China. One national cross-sectional study suggested that Tibetans had a significantly lower prevalence of diabetes than the majority Han population.¹⁴ Another national cross-sectional study found large ethnic differences in haemoglobin distribution and anaemia prevalence.¹⁵ To our knowledge, these are the only studies examining national ethnic variation in NCDs in China.

China's Southwest region is home to the 56 ethnic groups living in China.¹³ Ethnicity not only represents genetic diversity but also characterizes strong local identities for work and daily life, such as the Tibetan people in rural areas who usually live a nomadic life. Many ethnic minorities live in remote areas, speak different dialects and share some unique dietary habits, e.g. Tibetan people favouring butter tea¹⁶ or the Miao people's diet rich in sour soup (fermented acidic liquid).¹⁷ Although substantial efforts have been made to improve the socio-economic status of ethnic minorities,¹⁸ they still tend to be much poorer than their Han counterparts. Many ethnic minorities live in distinct geographical environments and climates. Tibetans, for example, live in the Tibetan Plateau with low barometric pressure, which could cause high-altitude hypoxia or decreased oxygen levels.¹⁹ In contrast, >100 million Han people live in the low-altitude Sichuan Basin and are exposed to severe air pollution.

To address the urgent need for data on the prevalence of, risk factors for and associated conditions of NCDs across various ethnicities in resource-constrained settings, the China Multi-Ethnic Cohort (CMEC) study was launched in 2017. This cohort study is supported by the National Key Research & Development Program of China. The CMEC aims to involve 100 000 middle-aged adults in five provinces of Southwest China. The recruitment started in May 2018 and was completed in September 2019. The study is based on a standardized survey approach and multiple stringent quality control (QC) measures. Such steps are essential, considering not only the substantial research gap but also that the study of NCDs across multiple ethnicities is likely to reveal important and potentially novel avoidable causes, which may inform disease prevention programmes in other populations worldwide.

Who is in the sample?

The CMEC was established in five provinces of Southwest China (Figure 1) and the baseline survey was conducted between May 2018 and September 2019. Given a full consideration of China's ethnic characteristics, population size and NCD patterns, we aimed to include participants from the Tibetan, Yi, Miao, Bai, Bouyei and Dong ethnic groups. In addition, to compare those ethnic minorities with the majority of people in China, the Han ethnic population was also recruited into our cohort. A multistage, stratified cluster sampling method was used to obtain samples from community-based populations. In the first stage, one to two minority settlements for each ethnic group were selected as our study sites. To achieve better representation of variation with regard to geography and development state, settlements from high plateaus, basins, rural areas

and highly air-polluted regions were given special consideration. In the second stage, 1–8 communities (depending on the size of communities) in each settlement were selected by the local Centres for Disease Control and Prevention (CDCs), taking into account migration status, local health conditions and, most importantly, ethnic structure. In the final stage, participants who met our inclusion criteria were invited to participate in our studies in consideration of both sex ratio and age ratio. The estimated population response rate was ~60% (60–90% in rural areas and 40–60% in urban areas).

The inclusion criteria included (i) age 30–79 years on the day of the investigation (except for Tibetan populations, whose inclusion criterion for age was 18–79 years, given the shorter life expectancy in this population); (ii) permanent residents, capability of completing baseline surveys and availability to complete follow-up studies; and (iii) complete questionnaire interviews, physical examinations and blood tests. The exclusion criteria were as follows: (i) inability to provide a unique national identification card; (ii) severe physical or mental diseases (e.g. schizophrenia and bipolar disorder); and (iii) refusal to comply with the requirements of the study. Overall, a total of 99 556 participants were enrolled in our cohort including 55 443 Han, 12 730 Tibetan, 6283 Yi, 5559 Miao, 6310 Bai, 5992 Bouyei and 7239 Dong ethnic people. All of the participants signed an informed consent form prior to data collection. Ethical approval was received from the Sichuan University Medical Ethical Review Board (K2016038).

How often will participants be followed up, and what is the rate of loss likely to be?

The follow-up stage will be conducted in both passive and active modes. In passive mode, we will conduct follow-up via linkage to the electronic records of disease and death registries, mainly relying on the National Electronic Disease Surveillance System (NEDSS) and the National Health Insurance (NHI) claim databases. Through annual data linkage with these established electronic disease registries, various health outcomes of the study participants will be obtained. The diagnoses of these conditions are based on well-accepted international standards. In China, the underlying and contributing causes of death are coded according to the 10th version of the International Statistical Classification of Diseases (ICD-10).²⁰ All deaths during follow-up will be checked regularly for cause-specific mortality through the death certificates reported to the regional CDCs and the death registries. In active mode, all surviving study participants will be invited to resurvey every 3–5 years. We will re-estimate the status of risk exposure and document the incidence of multiple diseases

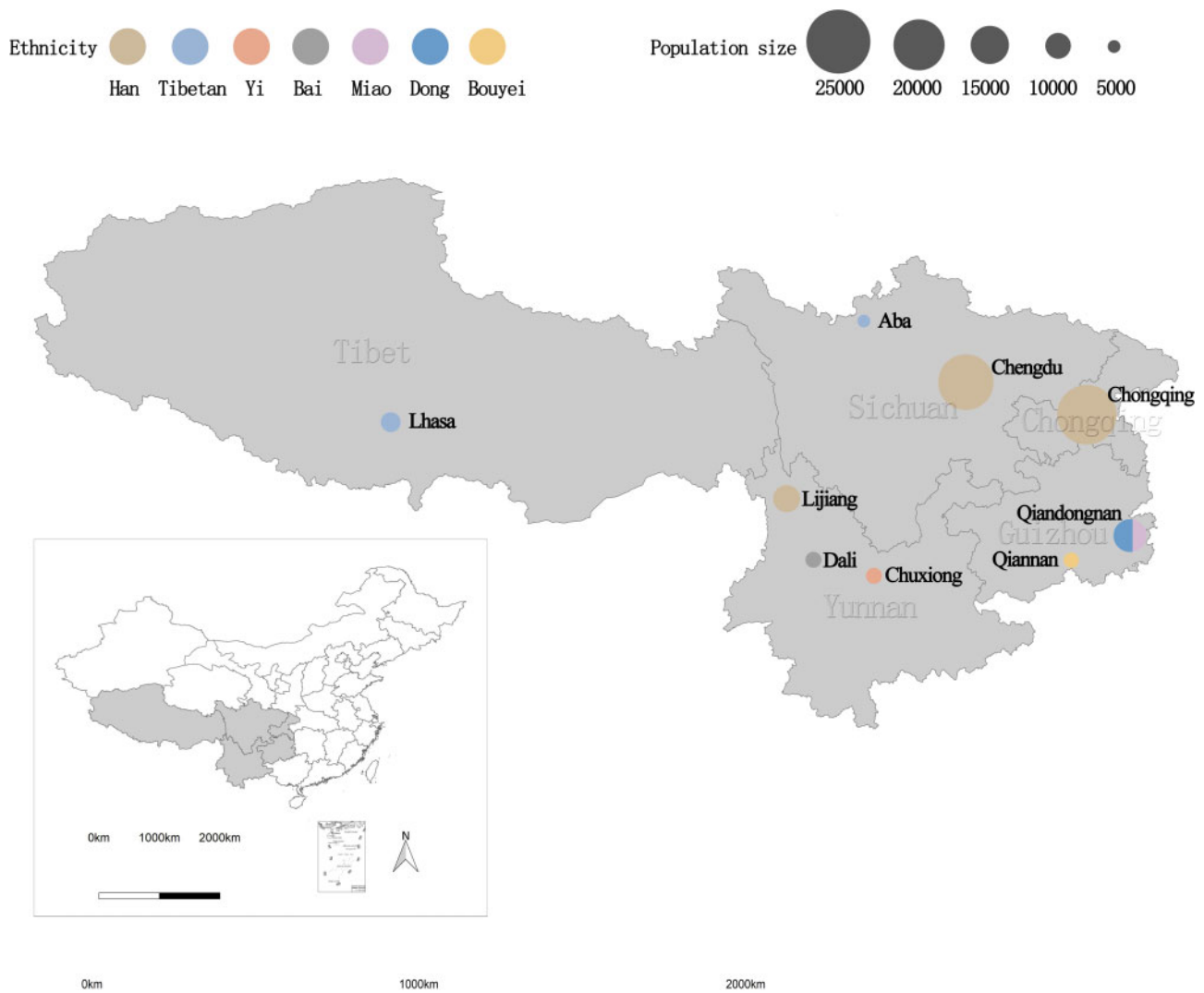


Figure 1 Spatial distribution of baseline populations

(e.g. cancer, stroke, heart attack, diabetes mellitus, hypertension) using the same measurements as in the baseline survey (see more details in the following section). We aim to minimize the expected rate of loss to less than an estimated 8% in the next round of follow-up.

What has been measured?

Table 1 summarizes the measurements collected from the CMEC baseline survey, which consisted of an electronic questionnaire with face-to-face interviews, medical examinations and clinical laboratory tests. In addition, we collected biological specimens, including blood, saliva and stool, which were stored for further use. In addition to the individual-level data for each participant, we also collected

regional-level data for each investigation site. More details on these measurements are described below.

Collaboration and field workflow

The CMEC baseline survey was conducted through the collaborative efforts of multiple agencies, including academic institutions, CDCs, local clinical centres, Third-Party Medical Laboratories (TPMLs) and local governments. Each agency has a clear role in the process of the investigation. Typically, the local government began the CMEC publicity campaigns a few weeks before the formal investigation. Meanwhile, the residents were fully informed of the benefits and requirements of participating in the CMEC study. Then, residents who were willing to participate made an appointment with the local clinical centre. The clinic visit for each participant typically took 90–

Table 1 Summary of measurements at the baseline survey in the China Multi-Ethnic Cohort study

Measurements	No. of variables	Variables
Electronic questionnaire		
Demographics and socio-economic status	21	National ID number, date of birth, sex, marital status, ethnic group, education, occupation, household income, household expenditure, type of medical insurance, contact information.
Smoking and indoor air pollution	20	Smoking status, smoking cessation, passive smoking exposure, pesticide exposure, indoor air pollution.
Alcohol consumption	11	Frequency of drinking, types of drink, amount of alcohol consumption, post-drinking symptoms.
Tea and other beverages	12	Frequency of drinking tea, types of tea, amount of tea consumption, frequency of drinking other beverages, types of beverage.
Health status	137	Quality of life (EQ-5D), self-rated health status, personal and family disease history.
Physical activity	24	Occupational physical activity, transportation physical activity, leisure-time physical activity, housework, weight change during last 12 months, weight in early adulthood.
Reproductive history (for women)	19	Age of menarche, menopause status, experience of pregnancy, history of breast feeding, history of contraceptive pills use, history of hysterectomy and of ovary/breast surgery, history of intrauterine contraceptive device, history of mammography.
Diet	139	Cooking oil consumption, flavouring, food frequency questionnaire (FFQ) ^a , nutritional supplements intake, history of severe food shortage, history of refrigerator usage, spicy food intake, Sichuan pepper intake.
Life events, sleep, psychological conditions and social support	28	Stress life events, sleep duration, sleep disturbance, napping, Patient Health Questionnaire (PHQ-2), Generalized Anxiety Disorder scale (GAD-2), social support and social capital.
Medical examination		
Fasting condition	1	Fasting time.
Medication history	2	Antihypertensive drugs, antiarrhythmic drugs.
Anthropometric measurement	4	Weight, height, hip circumference, waist circumference.
Physical examination	3	Fasting blood pressure (measured three times, OMRON HEM-8711), fasting heart rate (measured three times, OMRON HEM-8711), visual inspection (international standard visual acuity chart).
Lung function	1	Peak expiratory flow rate (KOKA PEF-3).
Osteopathic examination	1	Mineral density of anklebone (OSTEOKJ3000).
Electrocardiogram (ECG)	1	12-Lead ECG.
Abdominal ultrasonography	5 for male 8 for female	Liver, gallbladder, spleen, pancreas, kidney. Uterus, ovaries and fallopian tubes (additional for female participants).
Chest X-ray	2	Cardiovascular, lungs.
Chronic high-altitude disease ^b	7	Chronic high-altitude disease scale (seven questions or symptoms).
Clinical laboratory test		
Routine blood	24	White blood cells (WBC), red blood cells (RBC), haemoglobin (HGB), platelets (PLT), lymphocytes, monocytes, neutrophils, eosinophils.
Fasting blood glucose	1	Fasting blood glucose (FBG).
Lipid levels	4	Triglycerides (TG), cholesterol (CHOL), high-density lipoprotein cholesterol (HDL-CH) and low-density lipoprotein cholesterol (LDL-CH).
Hepatic function	5	Total protein (TP), albumin, globulin, alanine transaminase (ALT), aspartate amino transferase (AST), gamma-glutamyl transpeptidase (GGT).
Kidney function	3	Creatinine (Cr), urea, uric acid (UA).
Routine urine	12	Specific gravity, urobilinogen, urinary bilirubin, urinary ketone body, urine glucose, urinary protein, pH, urine nitrite.
Alkaline phosphatase	1	Alkaline phosphatase (ALP).
Creatine kinase-MB	1	Creatine kinase-MB (CK-MB).

(Continued)

Table 1 Continued

Measurements	No. of variables	Variables
Total bilirubin	1	Total bilirubin (TBIL).
Glycosylated haemoglobin	1	Glycated haemoglobin (HbA1C).
Biological specimens		
Blood sample	6	2 aliquots of serum (0.5 mL × 2), 2 aliquots of plasma (0.5 mL × 2), 1 aliquot of buffy coat (0.5 mL) and 1 aliquot of blood cells (0.5 mL).
Saliva	2	2 aliquots of saliva (1 mL × 2).
Stool	1	1 aliquot of stool (1 to ~2.5 g × 2).
Regional level data		
Hourly air pollution data	6	Sulfur dioxide, nitrogen dioxide, nitric oxide, fine particulate matter with particle size below 10 microns (PM10), fine particulate matter with particle size below 2.5 microns (PM2.5), ozone (China National Environmental Monitoring Centre).
Daily meteorological data	9	Air pressure, air temperature, precipitation, evaporation, relative humidity, wind direction and wind speed, sunshine hours and 0 cm ground temperature elements (National Meteorological Information Centre).
Socio-economic data	–	Specific variables selected according to research purposes (Yearbook data of national and provincial statistical departments).
Health related data	–	Specific variables selected according to research purposes (Yearbook data of national and provincial health departments).
Geographic data	–	Specific variables selected according to research purposes (National Earth System Science Data Sharing Platform).

^aIncludes common food items and ethnic group-specific food items.

^bOnly available for high-altitude areas.

120 min. The average daily recruitment rate was 70–80 participants per site.

Electronic questionnaire

The main content of our questionnaire referred to the study of the China Kadoorie Biobank as the prototype.²¹ Major modifications were made in sections regarding physical activity, habitual diet and psychological conditions to fully capture the unique features of the different ethnic minority groups. We used a tablet computer with a self-developed application (CMEC App) to collect the questionnaire information (see more details in [Table 1](#)). The information was collected by a face-to-face interview implemented by well-trained interviewers who were typically local college students with medical backgrounds. For a skilled interviewer, an average of 30–45 min was often required to complete a questionnaire.

The whole interview was audio recorded, and we documented the duration spent on each question. For data QC, on the same day of data collection, the data quality inspectors, who were chosen from excellent interviewers, drew random samples of questionnaires to assess their data quality by listening to the audio records. The sampling scheme was built on our computer system and could ensure that each interviewer was sampled at least once. Questionnaires

that were classified as unqualified were excluded from the final analysis.

Medical examinations

We conducted medical examinations mainly using the resources and personnel at local clinical centres. We implemented standardized training for the doctors and nurses before the investigation.

For those testing devices that the local clinical centres were rarely equipped with, we provided uniform devices for each site and trained the local staff to operate those devices, including bone mineral density densitometers (OSTEOKJ3000) and peak expiratory flowmeters (KOKA PEF-3). Particularly, for Tibet and high-altitude areas in Sichuan and Yunnan, qualified physicians used the chronic high-altitude disease scale, which consists of seven questions or symptoms for assessing chronic high-altitude disease and its severity.

The CDCs took the main responsibility for field QC, guided by a handbook and several forms to monitor key risk points. The main measures of QC included checking whether the measurement instruments were in good condition, inspecting whether the measurement was in accordance with standard operating procedures (SOPs) and randomly sampling participants to re-test their results.

Clinical laboratory tests and biobanking

All participants provided blood and urine samples on-site at the time of the baseline survey. Venous blood samples, collected after overnight fasting (at least 8 h), were used for clinical laboratory testing, including routine blood tests, fasting blood glucose, lipid levels and hepatic function. (Table 1). Mid-stream urine was collected for routine urine testing. A total of 4000 saliva and 1000 stool samples were collected from Tibetans and Han individuals residing in Sichuan Province for multi-omics analysis, such as the microbiome and systemic metabolism. All the samples were centralized in a biobank located at the West China School of Public Health, Sichuan University and stored at -80°C in cryogenic refrigerators. Computer systems for sample entry, access and temperature monitoring were developed and have been applied.

Blind panels of samples were periodically sent to the TPMLs for QC testing. Batches of controls were simultaneously put into the biobanking facility with samples and taken out for testing at regular intervals as a storage QC procedure.

Computer system and data cleaning

All of the above measurements were stored and managed in electronic form by a self-developed computer system, with the central server located at the West China School of

Public Health, Sichuan University. Different datasets were linked by the unique participant's ID. A visualized data organization schematic can be seen in Figure 2.

In addition to data storage and management, our computer system also had some built-in functions to help with data cleaning based on electronic data. The whole process can be divided into two parts: detection and verification. In the detection part, we checked for duplication, completeness, outliers and logic errors via predesigned algorithms. Once suspicious data errors were detected, we ran a further verification process by listening to the audio recordings, looking up examination reports or calling back the participants. Then, the data were revised, and the whole correction process was recorded by our computer system.

What has it found?

Table 2 shows the baseline characteristics of the participants. There were more women than men across the different ethnicities. The proportions of women ranged from 54.4% in the Han population in Basin (Chongqing, Chengdu) to 70.4% in the Bai population in the high plateau (Yunnan). Although we observed a similar age distribution and marital status across participants of different ethnicities, significant disparities were found in terms of highest education completed. For instance, 71.4% of Tibetans in Aba reported no formal schooling, whereas this number was as low as 10.7% among the Han in Basin.

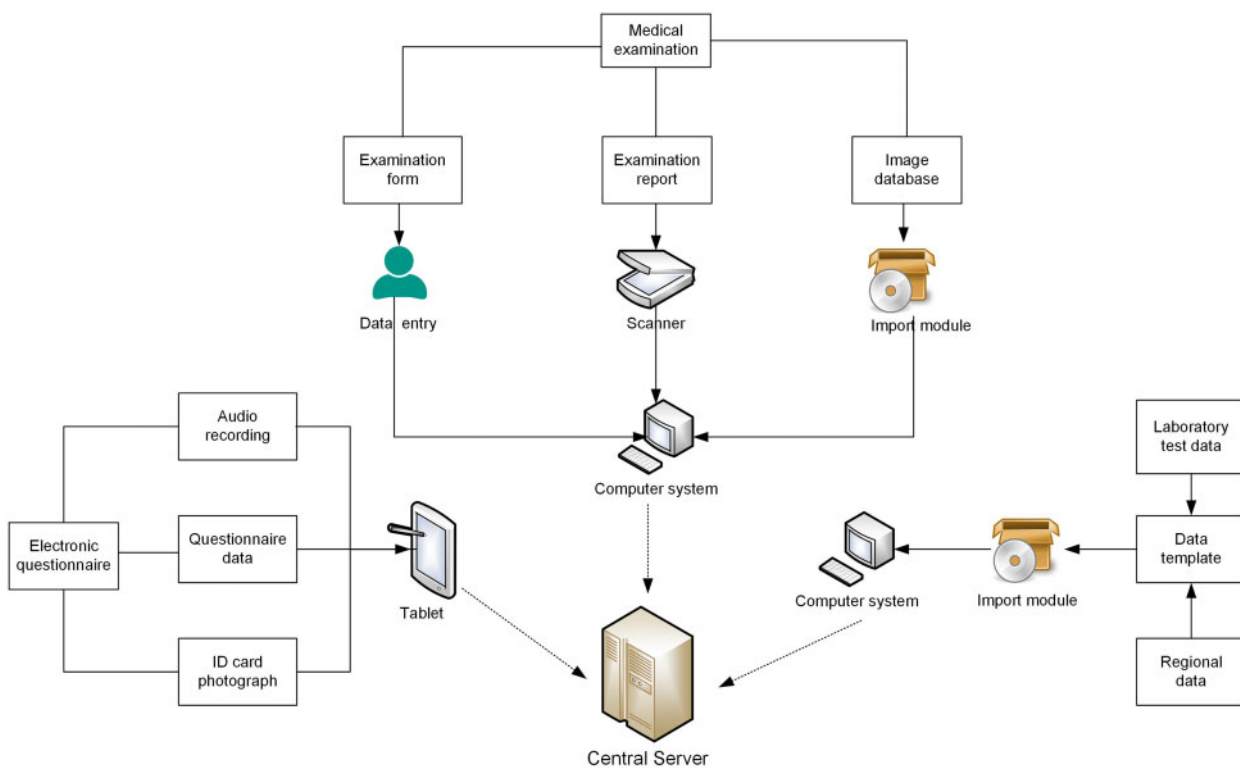


Figure 2 Schematic graph of the data collection and computer system

Table 2 Baseline characteristics of the participants in the China Multi-Ethnic Cohort

Characteristics of 99 556 participants	Han ethnic group in Basin (n = 44 900)	Han ethnic group in Yunnan (n = 10 543)	Yi ethnic group in Yunnan (n = 6283)	Bai ethnic group in Yunnan (n = 6310)	Tibetans in Aba (n = 4993)	Tibetans in Lhasa (n = 7737)	Dong ethnic group in Guizhou (n = 7239)	Bouyei ethnic group in Guizhou (n = 5992)	Miao ethnic group in Guizhou (n = 5559)
Gender, n (%)									
Male	20 478 (45.6)	3696 (35.1)	2088 (33.2)	1869 (29.6)	1927 (38.6)	3222 (41.6)	2594 (35.8)	1852 (30.9)	2067 (37.2)
Female	24 422 (54.4)	6847 (64.9)	4195 (66.8)	4441 (70.4)	3066 (61.4)	4515 (58.4)	4645 (64.2)	4140 (69.1)	3492 (62.8)
Ages (years), n (%) ^a									
<30	—	—	—	—	—	925 (12.0)	—	—	—
30–39	8382 (18.7)	1037 (9.8)	784 (12.5)	442 (7.0)	1268 (25.4)	1748 (22.6)	850 (11.7)	773 (12.9)	869 (15.6)
40–49	13 564 (30.2)	3112 (29.5)	1805 (28.7)	1759 (27.9)	1614 (32.3)	1870 (24.2)	1977 (27.3)	1783 (29.8)	1627 (29.3)
50–59	10 413 (23.2)	3605 (34.2)	2148 (34.2)	2124 (33.7)	1216 (24.4)	1934 (25.0)	2248 (31.1)	1897 (31.7)	1572 (28.3)
60–69	8499 (18.9)	2110 (20.0)	1157 (18.4)	1440 (22.8)	655 (13.1)	912 (11.8)	1484 (20.5)	1091 (18.2)	981 (17.7)
≥70	4042 (9.0)	679 (6.4)	389 (6.2)	545 (8.6)	240 (4.8)	348 (4.5)	680 (9.4)	448 (7.5)	510 (9.2)
Marital status, n (%)									
Married or cohabiting	39 893 (88.9)	9501 (90.1)	5516 (87.8)	5735 (90.9)	4358 (87.3)	6484 (83.8)	6377 (88.1)	5222 (87.1)	4817 (86.6)
Widowed	2466 (5.5)	190 (1.8)	127 (2.0)	106 (1.7)	224 (4.5)	154 (2.0)	180 (2.5)	214 (3.6)	200 (3.6)
Separated or divorced	2019 (4.5)	799 (7.6)	553 (8.8)	461 (7.3)	235 (4.7)	319 (4.1)	636 (8.8)	503 (8.4)	476 (8.6)
Never married	519 (1.2)	53 (0.5)	90 (1.4)	7 (0.1)	176 (3.5)	780 (10.1)	46 (0.6)	54 (0.9)	67 (1.2)
Highest education completed, n (%)									
No formal school	4787 (10.7)	2876 (27.3)	1569 (25.0)	2300 (36.5)	3565 (71.4)	3797 (49.1)	3603 (49.8)	2210 (36.9)	2454 (44.1)
Primary school	9709 (21.6)	3870 (36.7)	2625 (41.8)	2519 (39.9)	1058 (21.2)	2299 (29.7)	1021 (14.1)	1458 (24.3)	761 (13.7)
Middle school	14 499 (32.3)	2740 (26.0)	1671 (26.6)	1227 (19.5)	192 (3.9)	790 (10.2)	1211 (16.7)	1514 (25.3)	1083 (19.5)
High school	8073 (18.0)	653 (6.2)	279 (4.4)	199 (3.2)	53 (1.1)	405 (5.2)	596 (8.2)	462 (7.7)	584 (10.5)
College or university	7829 (17.4)	403 (3.8)	142 (2.3)	65 (1.0)	125 (2.5)	446 (5.8)	808 (11.2)	349 (5.8)	678 (12.2)
Annual income, whole family (Yuan/year), n (%)									
<12 000	4586 (10.2)	2261 (21.5)	1579 (25.1)	979 (15.6)	877 (17.6)	1771 (22.9)	2252 (31.2)	1968 (32.9)	1438 (25.9)
12 000–19 999	5435 (12.1)	2262 (21.5)	1798 (28.6)	1344 (21.4)	1693 (33.9)	2191 (28.3)	1515 (21.0)	1384 (23.1)	925 (16.7)
20 000–59 999	16 080 (35.9)	4570 (43.4)	2378 (37.9)	2652 (42.2)	2038 (40.8)	2627 (34.0)	1944 (26.9)	1784 (29.8)	1797 (32.4)
60 000–99 999	9480 (21.1)	805 (7.6)	345 (5.5)	726 (11.5)	246 (4.9)	630 (8.2)	794 (11.0)	519 (8.7)	824 (14.8)
≥100 000	9253 (20.6)	636 (6.0)	180 (2.9)	590 (9.4)	138 (2.8)	514 (6.7)	720 (10.0)	336 (5.6)	571 (10.3)
Smoking history, n (%)									
No smoking	31 650 (70.5)	7839 (74.4)	4941 (78.6)	4827 (76.5)	4556 (91.3)	5724 (74.0)	5679 (78.5)	4933 (82.3)	4322 (77.7)
Smoking ^b	10 273 (22.9)	2388 (22.7)	1169 (18.6)	1258 (20.0)	242 (4.9)	1651 (21.3)	1269 (17.5)	863 (14.4)	1030 (18.5)
Quit smoking ^c	2964 (6.6)	316 (3.0)	175 (2.8)	224 (3.6)	194 (3.9)	361 (4.7)	290 (4.0)	195 (3.3)	208 (3.7)
Drinking of alcohol, n (%)									
Never	21 161 (47.2)	7382 (70.0)	4442 (70.7)	5234 (83.1)	4647 (93.1)	5269 (68.1)	4125 (57.0)	2940 (49.1)	2505 (45.1)
Occasional ^d	18 373 (40.9)	1925 (18.3)	1270 (20.2)	669 (10.6)	314 (6.3)	2078 (26.9)	2469 (34.1)	2474 (41.3)	2635 (47.4)

(Continued)

Table 2 Continued

Characteristics of 99 556 participants	Han ethnic group in Basin (n = 44 900)	Han ethnic group in Yunnan (n = 10 543)	Yi ethnic group in Yunnan (n = 6283)	Bai ethnic group in Yunnan (n = 6310)	Tibetans in Aba (n = 4993)	Tibetans in Lhasa (n = 7737)	Dong ethnic group in Guizhou (n = 7239)	Bouyei ethnic group in Guizhou (n = 5992)	Miao ethnic group in Guizhou (n = 5559)
Regular ^c	5347 (11.9)	1236 (11.7)	572 (9.1)	395 (6.3)	31 (0.6)	389 (5.0)	643 (8.9)	577 (9.6)	420 (7.6)
Drinking of tea, n (%)									
Never	31 092 (69.3)	5804 (55.1)	5308 (84.5)	4406 (70.0)	1613 (32.3)	628 (8.1)	6228 (86.1)	5239 (87.5)	4815 (86.6)
Occasional ^f	2286 (5.1)	483 (4.6)	191 (3.0)	117 (1.9)	195 (3.9)	406 (5.3)	196 (2.7)	138 (2.3)	188 (3.4)
Regular ^g	11 503 (25.6)	4255 (40.4)	783 (12.5)	1775 (28.2)	3184 (63.8)	6700 (86.6)	814 (11.3)	614 (10.3)	557 (10.0)
Total physical activity (metabolic equivalent h/day), n (%) ^h									
<10	8043 (18.0)	877 (8.4)	435 (6.9)	839 (13.4)	1511 (30.5)	2157 (28.0)	993 (13.8)	909 (15.3)	851 (15.5)
10–19.9	15 790 (35.3)	1770 (16.9)	841 (13.4)	1581 (25.3)	1288 (26.0)	2096 (27.2)	1643 (22.8)	1406 (23.7)	1268 (23.0)
20–29.9	10 515 (23.5)	2164 (20.7)	1315 (21.0)	1339 (21.4)	755 (15.2)	1661 (21.6)	1587 (22.1)	1318 (22.2)	1162 (21.1)
30–39.9	5344 (12.0)	2148 (20.5)	1447 (23.1)	932 (14.9)	602 (12.2)	872 (11.3)	1259 (17.5)	947 (16.0)	892 (16.2)
≥40	5027 (11.2)	3522 (33.6)	2228 (35.6)	1570 (25.1)	800 (16.1)	923 (12.0)	1716 (23.8)	1358 (22.9)	1336 (24.3)
Frequency of physical exercise, n (%)									
Never	16 642 (37.1)	7267 (69.0)	4739 (75.4)	3516 (56.0)	2626 (52.8)	4288 (55.5)	5143 (71.1)	4105 (68.6)	4016 (72.3)
Occasional ⁱ	8381 (18.7)	716 (6.8)	513 (8.2)	927 (14.8)	1271 (25.6)	1710 (22.1)	609 (8.4)	498 (8.3)	466 (8.4)
Regular ^j	19 852 (44.2)	2552 (24.2)	1030 (16.4)	1837 (29.3)	1075 (21.6)	1734 (22.4)	1485 (20.5)	1382 (23.1)	1076 (19.4)
Reproductive history in women, n (%)									
Age at menarche <13 years	4218 (17.3)	880 (12.9)	233 (5.6)	493 (11.2)	159 (5.2)	141 (3.1)	380 (8.2)	331 (8.0)	312 (9.0)
Age at first live birth <20 years	740 (3.1)	325 (4.9)	610 (14.7)	163 (3.7)	892 (30.9)	390 (9.9)	277 (6.1)	202 (5.0)	300 (8.8)
≥3 Live births	2338 (9.7)	891 (13.2)	637 (15.3)	1151 (26.3)	1708 (59.0)	975 (24.7)	1923 (42.0)	1478 (36.2)	1212 (35.4)
Age at menopause <50 years	6070 (52.9)	1841 (51.6)	1075 (50.8)	1351 (55.9)	609 (54.9)	838 (57.3)	1648 (62.9)	1188 (56.2)	900 (52.9)
Ever used contraceptive pill	5265 (21.6)	464 (6.8)	603 (14.4)	173 (3.9)	120 (3.9)	621 (13.8)	797 (17.2)	631 (15.3)	721 (20.7)
High fat diet, n (%) ^k	31 409 (71.0)	8717 (83.0)	4185 (66.7)	5206 (84.3)	3502 (70.5)	4255 (55.3)	4582 (64.1)	3924 (67.1)	3281 (60.0)
High salt diet, n (%) ^l	12 227 (27.6)	4390 (41.8)	1826 (29.1)	1269 (20.6)	1256 (25.3)	3317 (43.1)	3662 (51.2)	2614 (44.7)	2785 (51.0)
Regular consumption of certain foodstuffs, n (%) ^m									
Rice	44 603 (99.4)	10 501 (99.8)	6267 (99.8)	6177 (99.9)	4606 (92.6)	6950 (90.1)	7037 (97.3)	5949 (99.7)	5411 (97.4)
Pasta	25 483 (56.8)	5652 (53.7)	1174 (18.7)	3195 (51.7)	2064 (41.5)	5137 (66.6)	2036 (28.1)	3531 (59.2)	2303 (41.5)
Coarse cereals	8686 (19.4)	1100 (10.5)	89 (1.4)	322 (5.2)	29 (0.6)	480 (6.2)	115 (1.6)	131 (2.2)	161 (2.9)
Tubers	12 706 (28.3)	5442 (51.7)	1780 (28.3)	4822 (78.0)	1537 (30.9)	4458 (57.8)	305 (4.2)	330 (5.5)	473 (8.5)
Meat/poultry	40 094 (89.4)	7419 (70.5)	3919 (62.4)	4396 (71.1)	4109 (82.6)	5566 (72.1)	6552 (90.6)	5345 (89.6)	4890 (88.1)
Fish/sea food	1918 (4.3)	380 (3.6)	64 (1.0)	321 (5.2)	NA ^r	36 (0.5)	238 (3.3)	97 (1.6)	94 (1.7)
Eggs	2520 (7.6)	4425 (42.1)	971 (15.5)	2316 (37.4)	551 (11.1)	1499 (19.4)	1321 (18.3)	2149 (36.0)	1326 (23.9)
Fresh vegetables	44 484 (99.2)	10 291 (97.8)	6212 (98.9)	6117 (98.9)	4582 (92.2)	7232 (93.7)	7166 (99.1)	5914 (99.1)	5498 (99.0)
Soybeans	7468 (16.7)	1445 (13.7)	519 (8.3)	1066 (17.2)	61 (1.2)	482 (6.3)	535 (7.4)	1189 (19.9)	596 (10.7)
Fresh fruits	31 865 (71.0)	8669 (82.4)	3423 (54.5)	4860 (78.6)	2698 (54.3)	3563 (46.2)	3894 (53.8)	3455 (57.9)	2453 (44.2)
Dairy products	17 012 (37.9)	1354 (12.9)	630 (10.1)	1070 (17.3)	2250 (45.3)	1845 (23.9)	391 (5.4)	610 (10.2)	470 (8.5)

(Continued)

Table 2 Continued

Characteristics of 99 556 participants	Han ethnic group in Basin (n = 44 900)	Han ethnic group in Yunnan (n = 10 543)	Yi ethnic group in Yunnan (n = 6283)	Bai ethnic group in Yunnan (n = 6310)	Tibetans in Aba (n = 4993)	Tibetans in Lhasa (n = 7737)	Dong ethnic group in Guizhou (n = 7239)	Bouyei ethnic group in Guizhou (n = 5992)	Miao ethnic group in Guizhou (n = 5559)
Spicy foods	28 429 (63.4)	8837 (84.0)	3233 (51.5)	4160 (67.3)	1850 (37.2)	2861 (37.1)	6738 (93.1)	5396 (90.5)	4899 (88.3)
Black pepper-seasoned foods	25 582 (57.0)	7677 (73.0)	2430 (38.7)	3306 (53.4)	1257 (25.3)	3205 (41.5)	1909 (26.4)	2612 (43.8)	2261 (40.7)
Sleeping status, n (%)									
Difficulty initiating or maintaining sleep ^a	17 300 (38.6)	3057 (29.1)	2225 (35.4)	1606 (26.3)	2236 (45.0)	1807 (23.4)	2694 (37.2)	2194 (36.8)	2107 (38.0)
Early morning awakening ^b	13 689 (30.5)	2747 (26.1)	2202 (35.1)	1440 (23.5)	2003 (40.3)	1511 (19.6)	2096 (29.0)	1752 (29.4)	1643 (29.6)
Taking sleeping pills ^c	1048 (2.3)	280 (2.7)	211 (3.4)	190 (3.1)	1305 (26.3)	164 (2.2)	75 (1.0)	115 (1.9)	104 (1.9)
Daytime dysfunction ^d	7430 (16.6)	1636 (15.6)	1250 (20.0)	489 (8.0)	1667 (33.5)	717 (9.3)	1024 (14.2)	979 (16.4)	966 (17.4)
Noon break	27 322 (60.9)	3574 (34.0)	1225 (19.5)	737 (12.1)	1329 (26.7)	2074 (26.9)	3525 (48.7)	3089 (51.8)	2378 (42.8)

^aThe inclusion criteria for age was 30–79 years on the day of investigation (except for the Tibetan population in Lhasa whose inclusion criteria for age was 18–79 years).

^bSmoking is defined as a total of >100 cigarettes smoked to date.

^cQuit smoking is defined as quitting for >6 months.

^dOccasional drinking of alcohol is defined as drinking on an average of ≤1–2 days per week in the past year.

^eRegular drinking of alcohol is defined as drinking on an average of ≥3–5 days per week in the past year.

^fOccasional drinking of tea is defined as drinking on an average of ≤1–2 days per week in the past year.

^gRegular drinking of tea is defined as drinking on an average of ≥3–5 days per week in the past year.

^hTotal physical activity includes job-related physical activity, transportation physical activity, leisure-time physical activity and housework.

ⁱOccasional physical exercise is defined as ≤1–2 times per week in the past year.

^jRegular physical exercise is defined as ≥3–5 times per week in the past year.

^kA high-fat diet is defined as an average daily oil intake >25 g.

^lA high salt diet is defined as an average daily salt intake >6 g.

^mRegular consumption of certain foodstuffs is defined as an average of ≥3 times per week in the past year.

ⁿDifficulty initiating or maintaining sleep is defined as at least 3 days per week in the past month.

^oEarly morning awakening is defined as at least 3 days per week in the past month.

^pTaking sleeping pills is defined as at least 1 day per week in the past month.

^qDaytime dysfunction is defined as at least 3 days per week in the past month.

Table 3 Physical examination results for participants in the China Multi-Ethnic Cohort

Characteristic	Han ethnic group in Basin	Han ethnic group in Yunnan	Yi ethnic group in Yunnan	Bai ethnic group in Yunnan	Tibetans in Aba	Tibetans in Lhasa	Dong ethnic group in Guizhou	Bouyei ethnic group in Guizhou	Miao ethnic group in Guizhou
Height, mean \pm SD, (cm)	158.5 \pm 8.3	158.3 \pm 7.5	157.1 \pm 7.7	159.3 \pm 7.7	158.6 \pm 8.3	161.0 \pm 8.6	155.4 \pm 7.8	155.1 \pm 7.5	154.4 \pm 7.9
Weight, mean \pm SD, (kg)	61.7 \pm 10.5	56.9 \pm 9.5	58.3 \pm 10.6	58.3 \pm 9.9	66.0 \pm 13.4	65.8 \pm 10.8	57.8 \pm 10.3	58.1 \pm 10.0	59.7 \pm 10.1
BMI ^a , <i>n</i> (%), (kg/m ²)									
<18.5	832 (1.9)	777 (7.7)	351 (5.7)	437 (7.1)	60 (1.3)	116 (1.9)	310 (4.7)	218 (3.9)	110 (2.2)
18.5 to <24	19 615 (44.5)	6117 (60.4)	3262 (53.1)	3545 (57.6)	1646 (34.2)	1699 (27.2)	3230 (48.5)	2641 (47.5)	1953 (38.2)
24 to <28	17 535 (39.8)	2684 (26.5)	1846 (30.1)	1753 (28.5)	1625 (33.7)	3400 (54.4)	2339 (35.2)	2006 (36.1)	2114 (41.4)
≥ 28	6096 (13.8)	555 (5.5)	684 (11.1)	420 (6.8)	1486 (30.9)	1030 (16.5)	776 (11.7)	695 (12.5)	935 (18.3)
Waist circumference, mean \pm SD, (cm)	82.7 \pm 9.5	77.3 \pm 9.3	80.2 \pm 10.2	82.8 \pm 9.7	86.9 \pm 12.2	91.6 \pm 11.9	83.0 \pm 10.1	82.6 \pm 9.5	83.6 \pm 9.9
Hipline, mean \pm SD, (cm)	93.8 \pm 6.7	89.4 \pm 6.1	91.9 \pm 6.7	93.0 \pm 6.3	97.6 \pm 9.1	100.1 \pm 8.9	91.0 \pm 6.3	92.6 \pm 6.3	92.3 \pm 6.4
Heart rate, mean \pm SD, (/min) ^b	77.2 \pm 10.6	75.5 \pm 11.7	73.1 \pm 11.7	74.8 \pm 11.4	71.9 \pm 11.3	74.3 \pm 10.9	74.7 \pm 10.7	75.1 \pm 10.7	73.8 \pm 11.1
SBP, <i>n</i> (%), (mmHg) ^b									
<120	16 628 (37.7)	5208 (51.4)	2126 (34.6)	2290 (37.2)	1889 (39.3)	3720 (59.7)	3009 (45.3)	2538 (45.8)	2198 (43.2)
120–139	16 938 (38.4)	3308 (32.6)	2274 (37.0)	2298 (37.3)	2045 (42.6)	1698 (27.3)	2282 (34.3)	1778 (32.1)	1763 (34.6)
≥ 140	10 533 (23.9)	1620 (16.0)	1745 (28.4)	1574 (25.5)	872 (18.1)	812 (13.0)	1358 (20.4)	1228 (22.2)	1128 (22.2)
DBP, <i>n</i> (%), (mmHg) ^b									
<80	24 697 (56.0)	5671 (55.9)	3149 (51.3)	3414 (55.4)	2726 (56.8)	3464 (55.7)	3476 (52.3)	2579 (46.5)	2450 (48.1)
80–89	12 555 (28.5)	2771 (27.3)	1719 (28.0)	1721 (27.9)	1277 (26.6)	1637 (26.3)	1894 (28.5)	1720 (31.0)	1497 (29.4)
≥ 90	6843 (15.5)	1695 (16.7)	1276 (20.8)	1027 (16.7)	796 (16.6)	1124 (18.1)	1283 (19.3)	1253 (22.6)	1146 (22.5)
Pulmonary function, mean \pm SD, (L/min)	355.8 \pm 125.1	349.5 \pm 100.6	328.9 \pm 104.3	286.6 \pm 89.1	292.4 \pm 132.6	200.6 \pm 94.8	361.0 \pm 94.3	348.9 \pm 84.8	366.4 \pm 97.0
Bone density, <i>n</i> (%)									
Normal	23 976 (54.4)	6421 (63.3)	4330 (70.5)	3331 (54.1)	2922 (60.7)	4043 (64.8)	3995 (60.0)	3503 (62.9)	3035 (59.3)
Osteopaenia	17 082 (38.7)	3479 (34.3)	1683 (27.4)	2490 (40.4)	1183 (24.6)	1686 (27.0)	2435 (36.6)	1844 (33.1)	1893 (37.0)
Osteoporosis	3034 (6.9)	240 (2.4)	133 (2.2)	341 (5.5)	706 (14.7)	506 (8.1)	231 (3.5)	227 (4.1)	190 (3.7)

^aBMI, body mass index.^bHeart rate, SBP and DBP were measured three times for each participant and the statistics are based on the average of the three measurements.

Table 4 Biochemical characteristics of participants in the China Multi-Ethnic Cohort (mean ± standard deviation)

Biochemical index ^a	Han ethnic group in Basin	Han ethnic group in Yunnan	Yi ethnic group in Yunnan	Bai ethnic group in Yunnan	Tibetans in Aba	Tibetans in Lhasa	Dong ethnic group in Guizhou	Bouyei ethnic group in Guizhou	Miao ethnic group in Guizhou
FBG (mmol/l)	5.5±1.5	5.2±1.2	5.6±1.4	5.5±1.2	4.6±1.1	4.5±1.3	5.6±1.4	5.4±1.4	5.5±1.4
ALP (U/l)	86.7±29.4	84.1±27.7	84.0±29.0	90.2±28.5	114.7±43.5	113.2±37.9	92.6±32.7	74.3±23.8	81.2±26.9
TBIL (μmol/l)	12.7±7.1	11.5±5.7	10.2±5.1	12.0±6.3	12.7±8.5	13.9±7.0	9.9±5.1	10.1±4.6	10.3±5.4
HbA1C (%)	5.7±0.9	5.8±0.8	5.7±0.8	5.8±0.8	5.9±0.8	6.0±1.0	5.7±1.0	5.8±1.0	5.7±1.0
WBC (10 ⁹ /l)	5.9±1.7	6.6±2.0	6.4±1.8	6.4±1.7	5.7±1.6	6.2±1.7	6.6±1.8	6.2±1.7	6.5±1.8
RBC (10 ¹² /l)	4.8±0.6	5.0±0.5	4.8±0.5	5.0±0.5	5.3±0.6	5.1±0.8	4.7±0.6	4.7±0.6	4.8±0.6
HGB (g/l)	144.2±16.4	150.6±15.1	147.8±14.7	151.6±15.5	160.3±21.2	154.5±28.5	140.8	137.4	143.6
PLT (10 ⁹ /l)	192.9±67.8	217.1±59.4	232.6±63.1	203.4±62.1	240.2±64.6	233.3±88.3	217.6	227.0	239.2
TG (mmol/l)	1.7±1.4	2.0±1.7	1.9±2.1	1.6±1.3	1.0±0.6	1.3±0.8	2.0±1.8	1.8±1.6	1.9±1.9
CHOL (mmol/l)	5.0±1.0	5.0±1.0	5.5±1.1	5.1±1.0	4.8±1.0	4.7±1.0	4.9±1.0	5.0±1.0	5.0±1.0
HDL-CH (mmol/l)	1.5±0.4	1.5±0.4	1.6±0.4	1.7±0.4	1.3±0.3	1.3±0.3	1.5±0.4	1.5±0.3	1.5±0.3
LDL-CH (mmol/l)	2.9±0.8	3.0±0.8	3.5±1.0	3.0±0.8	3.1±0.8	2.8±0.7	3.0±0.9	2.5±0.7	2.9±0.9
TP (g/l)	76.3±4.6	75.6±4.6	76.2±4.8	74.0±4.3	77.6±5.1	76.5±5.3	79.2±4.6	77.5±4.5	78.6±5.4
ALT (U/l)	24.2±19.0	20.5±16.5	21.4±16.0	23.3±15.8	30.8±47.7	32.3±26.2	24.9±21.6	24.5±20.4	23.9±19.4
AST (U/l)	26.1±13.5	25.6±12.5	25.4±14.6	28.0±13.2	28.4±24.5	29.4±18.2	28.6±18.5	28.0±17.7	27.7±15.6
GGT (U/l)	34.2±49.6	39.7±63.1	47.2±92.0	38.8±54.2	42.3±46.6	52.4±90.5	43.9±81.0	38.8±53.1	41.1±68.5
Cr (μmol/l)	67.9±18.4	78.3±20.4	76.6±20.8	72.1±18.9	65.8±16.7	66.6±15.0	70.4±29.7	64.3±24.3	66.8±24.8
Urea (mmol/l)	5.5±1.5	5.1±1.5	5.3±1.6	5.2±1.5	5.2±1.5	4.7±1.4	5.2±1.7	5.3±1.6	5.1±1.7
UA (μmol/l)	322.3±86.7	318.6±87.7	299.8±87.7	286.1±83.1	316.7±89.8	334.3±89.3	337.6	315.7	336.0
CK-MB (U/l)	16.7±13.8	-	-	-	16.7±9.1	-	-	-	-

^aFBG, fasting blood glucose; ALP, alkaline phosphatase; TBIL, total bilirubin; HbA1C, glycated haemoglobin; WBC, white blood cells; RBC, red blood cells; HGB, haemoglobin; PLT, platelets; TG, triglycerides; CHOL, cholesterol; HDL-CH, high-density lipoprotein cholesterol; LDL-CH, low-density lipoprotein cholesterol; TP, total protein; ALT, alanine transaminase; AST, aspartate amino transferase; GGT, gamma-glutamyl transpeptidase; Cr, Creatinine; UA, uric acid; CK-MB, creatine kinase-MB.

The proportions of participants who never smoked (91.3%) or drank alcohol (93.1%) were relatively high among Tibetans in Aba compared with other populations in our study. Additionally, we noted variations regarding some lifestyle variables within an ethnic group, e.g. the proportion that never consumed tea was 32.3% for Tibetans in Aba and 8.1% for Tibetans in Lhasa.

Table 3 presents the characteristics of the participants' medical examinations. The mean height was between 154.4 cm in Miao peoples and 161.0 cm in Tibetans from Lhasa. The mean weight was between 56.9 kg in the Han population in the high plateau (Yunnan) and 66.0 kg in Tibetans from Aba. The proportion of subjects who had systolic blood pressure (SBP) >140 mmHg was low (13.0%) in Lhasa Tibetans, and the proportion who had diastolic blood pressure (DBP) >90 mmHg was low (15.5%) in Basin Han individuals. In terms of bone density, osteoporosis was prevalent in Aba Tibetans (14.7%) compared with other populations.

Table 4 indicates the biochemical characteristics of the participants measured at baseline. The mean fasting blood glucose (FBG) was significantly low in Tibetans from Lhasa (4.5 mmol/L) and Tibetans from Aba (4.6 mmol/L). The mean glycated haemoglobin (HbA1C) was between 5.7 and 6.0% across different ethnicities. With regard to triglycerides (TG), the highest TG (2.0 mmol/L) was measured among the Han ethnic group in the high plateau (Yunnan), whereas the lowest TG (1.0 mmol/L) was found among Tibetans in the high plateau of Aba.

Figure 3 presents the prevalence of self-reported diseases, which demonstrates the diverse disease patterns among different ethnic groups. The highest prevalence of hypertension was observed in the Bai (20.5%) and Yi

(20.2%) ethnic groups in the high plateau, whereas the lowest was reported in Tibetans from Aba (11.4%). Diabetes was mostly prevalent among the Han ethnic group in Basin (6.5%). Regarding coronary heart disease, substantial disparity was observed. The most prevalent group was Tibetans from Lhasa (3.7%), whereas the least prevalent group was their rural counterpart, Tibetans from Aba (0.2%). Overall, cancer was not frequently reported, with the highest observation in the Bouyei ethnic group in Guizhou (1.4%).

What are the main strengths and weaknesses?

This study has some limitations worth mentioning. First, we relied on self-reported data for some variables, such as lifestyle factors, which may incur recall bias. Second, we excluded populations aged outside of 30–79 years (except for the Tibetan populations in which 18–30-year-olds were also included), which may have ignored information on early life exposures. Despite the above limitations, our study is unique. First, to our knowledge, this is the first large-scale cohort study focusing on ethnic minority groups in China. Due to the substantial disparity in terms of social and economic development among the ethnic minority groups, as well as their distinct and diverse exposures and outcomes, the findings based on our cohort study have the potential to not only add evidence to the extant body of literature on the prevalence and risk factors of certain NCDs but also reveal novel knowledge that may serve as a reference to other global populations. Second, our study involved multi-ethnic individuals, involving mixed groups of urban and rural residents, plateau and basin inhabitants, populations living in highly concentrated and air polluted

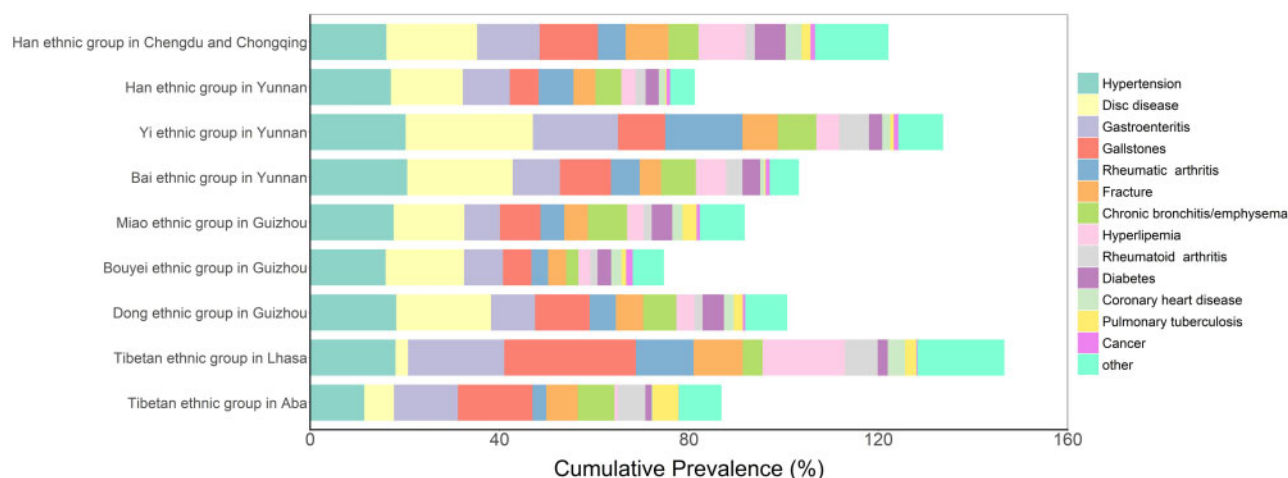


Figure 3 Prevalence of some selected self-reported diseases among the participants in different ethnic groups. (Cumulative prevalence adds together prevalences of all individual chronic diseases, so it may exceed 100%)

areas and those in remote mountainous areas. This would allow one to conduct comparative studies of the effects of various environmental exposures on people's health outcomes from different angles. For instance, Tibetan populations in rural areas (Aba prefecture) live a nomadic life, which is not shared by their urban counterparts (those in Lhasa). One will thus be able to investigate significant disparities of lifestyle even within the same ethnic group. Third, our baseline survey collected detailed and comprehensive information. For instance, in addition to demographics, socio-economics, lifestyles and health-related histories, we also collected biological specimens and performed medical examinations and clinical laboratory tests for every participant. Fourth, our study is funded by the National Key Research and Development Program. As the highest-level research programme in China, this study is strongly supported by the different levels of government, CDCs and local clinic centres. Therefore, related resources, field surveys, public cooperation and electronic medical records checks can be guaranteed and thus may improve the data quality and minimize the loss of participants at follow-up. Fifth, follow-up information will be obtained through annual data linkage to electronic records of disease and death registries, as well as re-interviews every 3–5 years. This follow-up strategy enables accurate and reliable tracking of both exposures and outcomes. Finally, various methods have been rigorously implemented at every step to ensure the quality of the data.

Profile in a nutshell

- The China Multi-Ethnic Cohort (CMEC) is a community population-based prospective observational study aiming to address the urgent need for understanding NCDs prevalence, risk factors and associated conditions in resource-constrained settings for ethnic minorities in China.
- A total of 99 556 participants aged 30–79 years (Tibetan populations include those aged 18–30 years) from the Tibetan, Yi, Miao, Bai, Bouyei, Dong and Han ethnic groups in Southwest China were recruited between May 2018 and September 2019.
- All surviving study participants will be invited for re-interview every 3–5 years with concise questionnaires to review risk exposures and disease incidence. Furthermore, the vital status of study participants will be followed-up through linkage with established electronic disease registries annually.
- The CMEC baseline survey collected data with an electronic questionnaire and face-to-face interviews, medical examinations and clinical laboratory tests.

Furthermore, we collected biological specimens, including blood, saliva and stool, for long-term storage. In addition to the individual-level data, we also collected regional-level data for each investigation site.

Collaborations are welcome. Please send specific ideas to the corresponding author at: xingzhao@scu.edu.cn.

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

To maximize the use of the CMEC data, we welcome collaboration from all over the world. Currently, the database is not accessible to be downloaded publicly because of some sensitive information. However, researchers interested in our study should contact the corresponding author at xingzhao@scu.edu.cn by providing specific ideas.

Funding

This work was supported by the National Key R&D Program of China (Grant no: 2017YFC0907305, 2017YFC0907300).

Acknowledgements

The most important acknowledgement is to all the participants in the study, as well as to all the interviewers and members of the survey teams in each of the study sites and communities. We also thank Drs Carine Ronsmans, Peng Jia and Huan Song for their critical reading of the manuscript.

Conflict of Interest

None declared.

References

1. Gowshall M, Taylor-Robinson SD. The increasing prevalence of non-communicable diseases in low-middle income countries: the view from Malawi. *Int J Gen Med* 2018;11:255–64.
2. Collaborators NC. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *Lancet* 2018;392:1072–88.
3. Yf W, Lim H, Wu Y. Growing global burden of chronic noncommunicable diseases and an alarming situation in China. *Beijing Da Xue Xue Bao Yi Xue Ban* 2012;44:688–93.
4. Rembert N, He K, Judd SE, McClure LA. The geographic distribution of trace elements in the environment: the REGARDS study. *Environ Monit Assess* 2017;189:84.
5. Wu F, Guo Y, Chatterji S *et al.* Common risk factors for chronic non-communicable diseases among older adults in China, Ghana, Mexico, India, Russia and South Africa, The Study on

- Global AGEing and Adult Health (SAGE) wave 1. *BMC Public Health* 2015;15:88.
6. 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.
 7. Liu X, Mao Z, Li Y *et al.* Cohort Profile: The Henan Rural Cohort: a prospective study of chronic non-communicable diseases. *Int J Epidemiol* 2019;48:1756–6j.
 8. Wang X, Lu M, Qian J *et al.* Rationales, design and recruitment of the Taizhou Longitudinal Study. *BMC Public Health* 2009;9: 223.
 9. Perini W, Snijder MB, Peters RJG, Kunst AE. Ethnic disparities in estimated cardiovascular disease risk in Amsterdam, the Netherlands: The HELIUS study. *Neth Heart J* 2018;26:252–62.
 10. Fang CY, Tseng M. Ethnic density and cancer: a review of the evidence. *Cancer* 2018;124:1877–903.
 11. Zhu Y, Sidell MA, Arterburn D *et al.* Racial/ethnic disparities in the prevalence of diabetes and prediabetes by BMI: Patient Outcomes Research To Advance Learning (PORTAL) Multisite Cohort of Adults in the U.S. *Dia Care* 2019;42:2211–19.
 12. Aarab R, Vijverberg SJH, Prins M *et al.* Prevalence of and factors associated with adult-onset asthma in different ethnic groups: The HELIUS study. *Respir Med* 2019;150:113–19.
 13. National Bureau of Statistics of China. *Tabulation on the 2010 Population Census of the People's Republic of China*. <http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm> (10 February 2020, date last accessed).
 14. Wang L, Gao P, Zhang M *et al.* Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. *Jama* 2017;317: 2515–23.
 15. Zhang X, He Y, Xie X, Ji M, Ma X, Yu Z. Distribution of hemoglobin and prevalence of anemia in 10 ethnic minorities in China. *Medicine* 2017;96:e9286.
 16. Xu S, Jiayong Z, Li B *et al.* Prevalence and clustering of cardiovascular disease risk factors among Tibetan adults in China: a population-based study. *PLoS One* 2015;10: e0129966.
 17. Qingli T. Study on characteristic components of fermented acidic liquid from miao-nationality in Guizhou province [J]. *Sci Technol Food Ind* 2005;9: 165–6.
 18. Bao Y, Sirijimodege. Suggestions on promoting the economic development of Chinese minorities in the new era. *J Fin Econ Theory* 2019;5:9.
 19. West JB. High-altitude medicine. *Am J Respir Crit Care Med* 2012;186:1229–37.
 20. World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision*. Geneva: WHO, 2004.
 21. Chen Z, Chen J, Collins R *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;40:1652–66.