

## Cohort Profile

# Cohort Profile: Regional Ethnic Cohort Study in Northwest China

**Chao Li,<sup>1†</sup> Yezhou Liu,<sup>1†</sup> Guoshuai Shi,<sup>1</sup> Baibing Mi,<sup>1</sup> Lingxia Zeng,<sup>1</sup> Qiang Li,<sup>1</sup> Yuan Shen,<sup>1</sup> Yaling Zhao,<sup>1</sup> Leilei Pei,<sup>1</sup> Yijun Kang,<sup>1</sup> Shengbing Xiao,<sup>1</sup> Xin Liu ,<sup>1</sup> Jianghong Dai,<sup>2</sup> Yuhong Zhang,<sup>3</sup> Xinhua Wang,<sup>4</sup> Xiaojie Wang,<sup>5</sup> Duolao Wang,<sup>6</sup> Hong Yan<sup>1\*</sup> and Shaonong Dang<sup>1\*</sup>**

<sup>1</sup>Department of Epidemiology and Biostatistics, School of Public Health, Xi'an Jiaotong University Health Science Center, Shaanxi, Xi'an, China, <sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang Medical University, Urumqi, Xinjiang, China, <sup>3</sup>Department of Epidemiology and Biostatistics, School of Public Health, Ningxia Medical University, Yinchuan, Ningxia, China, <sup>4</sup>Department of Environmental and Occupational Medicine, School of Public Health, Gansu University of Chinese Medicine, Lanzhou, Gansu, China, <sup>5</sup>Department of Chronic and Noncommunicable Disease Control and Prevention, Qinghai Center of Disease Prevention and Control, Xining, Qinghai, China and <sup>6</sup>Department of Clinical Science, Liverpool School of Tropical Medicine, Liverpool, UK

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author. School of Public Health, Xi'an Jiaotong University Health Science Center, No. 76, Yanta West Road, Xi'an, Shaanxi, CN 710061, China. E-mail: tjdsnh@xjtu.edu.cn and School of Public Health, Xi'an Jiaotong University Health Science Center, No. 76, Yanta West Road, Xi'an, Shaanxi, CN 710061, China. E-mail: yanhonge@mail.xjtu.edu.cn

Received 20 March 2021; Editorial decision 2 September 2021; Accepted 14 September 2021

## Why was the cohort set up?

Non-communicable diseases (NCDs) contribute to more than two-thirds of deaths worldwide. Approximately 80% of NCDs occur in low- and middle-income countries, and 33% of the deaths from NCDs are in people younger than 60 years old.<sup>1–4</sup> Although great efforts have been made to control NCDs in China, their prevalence continues to increase.<sup>5–8</sup> The prevalence of NCDs in China reportedly increased from 15.3% in 2003 to 29.5% in 2013.<sup>9</sup> The burden of NCDs and associated healthcare costs is rising for older people with chronic diseases who require long-term medical services and support. Patients with NCDs face the risk of disability and life-threatening conditions that result from complications of NCDs because of a lack of medical-service resources and coverage by healthcare systems.<sup>5,9</sup>

The region of Northwest China (31°–50°N, 73°–111°E) is considered to be relatively poorer (in financial terms) than the central and eastern provinces. It includes the provinces of Shaanxi, Gansu and Qinghai, and the autonomous regions of Ningxia Hui and Xinjiang Uygur.<sup>10</sup> The topography of the region is highly varied and includes large stretches of arid desert, wastelands, fertile oases, grassy plateaus and high mountain ranges. These provinces face various environmental issues because of decreased water availability and an arid climate.<sup>11,12</sup> Northwest China is also one of the most ethnically diverse areas, with 17 ethnic minorities and two of the five autonomous regions that are designated minority areas. Lifestyle and dietary habits differ between ethnic groups, and high rates of smoking and drinking in many ethnic populations further increase the risk of NCDs.<sup>13–17</sup>

### Key Features

- The Regional Ethnic Cohort Study in Northwest China (RECS)—a community, population-based, prospective observational study—aims to investigate specific aetiological causes and long-term health hazards of non-communicable diseases among different ethnic groups in the north-west region of China.
- A total of 118 572 participants aged 35–74 years from five provinces of Northwest China, including people from the Han, Uygur, Kazakh, Hui and Tibetan ethnic groups, were recruited between June 2018 and May 2019.
- Approximately 10% of the participants will be actively followed every 3 years via face-to-face interviews, using concise questionnaires to review risk exposures and disease incidence, and with the collection of biological specimens, including blood, saliva and stool samples. Passive follow-up will be conducted by periodic linking (every 6 months) of baseline survey data to established electronic disease registries.
- The questionnaire survey, regular medical examination and the collection and storage of blood samples were conducted in the RECS baseline survey for all participants. Approximately 40% of the participants had blood and urine samples assessed for several biochemical indices.
- Overweight/obesity and hypertension have become a prominent public health problem in Northwest China, especially in the Uygur ethnic group. Low prevalences of self-reported diabetes and cardiovascular disease among several ethnic groups (other than Han ethnicity) indicate lower levels of access to health services and chronic-disease management among the regional ethnic minorities of Northwest China.
- Collaborations from all over the world are welcome. Please contact the corresponding author at [tjdshn@xjtu.edu.cn](mailto:tjdshn@xjtu.edu.cn) with specific research ideas or questions.

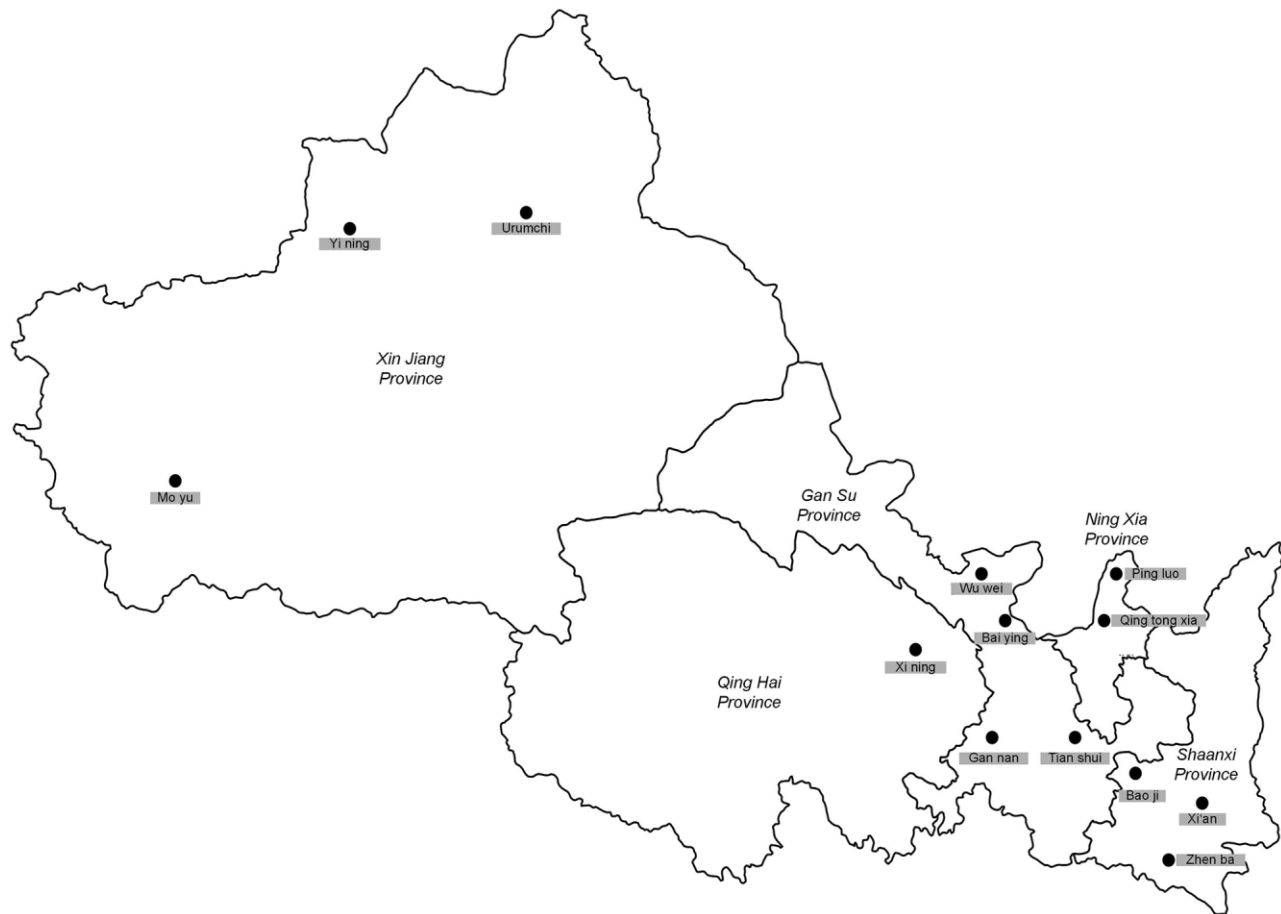
China has experienced tremendous changes in the past 30 years, during which time its rapid economic development has substantially influenced environmental exposure (severe air pollution), economic inequalities among regions and ethnicities, and diet and lifestyle changes in the Chinese population. These changes are also thought to have contributed to NCD development and mortality.<sup>18–20</sup> To address the urgent need for data on the prevalence of and risk factors for NCDs and associated conditions in resource-constrained settings, the Regional Ethnic Cohort Study in Northwest China (RECS), supported by the National Key R&D Program of China (NKRP), aimed to establish a nationally representative cohort, including people from different ethnicities and areas. The RECS, as one important part of the NKRP, will provide a reliable source of data for studies investigating specific aetiological causes and long-term health hazards of NCDs among different ethnic groups in Northwest China. The findings from studies using the cohort data will serve to promote the health conditions of the local population and to provide references for other countries to formulate prevention and control strategies for NCDs.

### Who is in the cohort?

This cohort consists of people in 13 areas of five western provinces of Northwest China (Xinjiang: Moyu county, Yining city, Urumchi city; Gansu: Wuwei city, Tianshui city, Gannan city, Baiying city; Qinghai: Xining city; Ningxia:

Qingtongxia city, Pingluo city; Shaanxi: Zhenba county, Baoji city, Xi'an city). The areas were chosen according to local disease patterns, exposures to certain risk factors, distribution of ethnic groups, population size and stability, quality of death and disease registries, and local commitment and capacity (Figure 1). The cohort was planned to include 110 000 participants and is managed by each provincial project group (Xinjiang: Xinjiang Medical University; Gansu: Gansu University of Chinese Medicine; Qinghai: Qinghai Center of Disease Prevention and Control; Ningxia: Ningxia Medical University; Shaanxi: Xi'an Jiaotong University). Each group is responsible for the baseline survey measurements, including the questionnaire interview, blood-sample collection and storage, and further follow-up work. The Shaanxi project team operates as the general manager in charge of the project. The baseline survey was conducted between June 2018 and May 2019.

After selecting the sites for inclusion in the cohort, the local community health centres and township hospitals advertised the study to the residents they served to recruit the participants. Permanent residents who met the inclusion criteria were not formally invited but could volunteer to participate, to help ensure the compliance of participants. After the target sample size for each village/community was reached, the local survey ended. A neighbouring village/community was used if the sample size in the first village/community was not reached. Participants were asked to show their national identity cards before entering the cohort, and only those whose address on the identity card



**Figure 1** Spatial distribution of participants in the baseline survey of the Regional Ethnic Cohort Study in Northwest China

showed them to be a resident of a cohort site were included, as we were unable to follow up people who were not permanent local residents.

The selection of sites was based on the population size and ethnic distribution of each province.<sup>21</sup> Therefore, the target sample sizes were 40 000, 30 000, 15 000, 20 000 and 5000 in Shaanxi (the largest proportion of the population), Xinjiang (multi-ethnic gathering area), Ningxia (gathering area of Hui ethnic), Gansu and Qinghai provinces, respectively. As the China Multi-Ethnic Cohort, which is also part of the NKRK, investigated Tibetans in Tibet, we reduced the sample size in Qinghai province and focused on Han ethnic participants living on the plateau, and enlarged the sample size in Shaanxi province to ensure that we reached the total target sample size of 110 000.

A baseline survey was conducted via face-to-face interviews and clinic visits, with assistance from local health staff. The participants in the cohort consisted of permanent residents aged 35–74 years, but any individuals attending the baseline survey who were slightly outside the target age

range were not turned away, to encourage participation. We excluded people who had severe mental disease or disabilities that hindered their ability to communicate, so we could obtain accurate data. All participants gave written informed consent, which allowed us to access their medical records and long-term storage supplies of blood samples, solely for medical-research purposes.

### How often will they be followed up?

The National Death Surveillance System (NDSS), Chronic Disease Registries (CDR), National Central Cancer Registry (NCCR) and National Health Insurance Datasets (NHID) enable us to conduct passive follow-up by periodic linking (every 6 months) of the baseline survey data to each of these data sets. The NDSS and CDR are both based on China's disease-surveillance points system. The causes of death of cohort participants are regularly monitored through the use of official death certificates that are reported to the regional Centers for Disease Control and

Prevention (CDC) where the NDSS is based (the NDSS is available in all 13 study areas). Information on chronic diseases (e.g. diabetes, hypertension, heart failure, stroke and cancer) is being collected through linkage with the NCCR, CDR and NHID. The new NHID are now fully established in all 13 study areas, and future information on the incidence of chronic disease and hospitalization will be primarily extracted from these data sets.<sup>22–26</sup> As the NCCR and CDR are still under construction, their coverage of the sites is limited, so they are not primary data sources but are used as complementary sources for passive follow-up. The unique national identity card and health-insurance card numbers were used as key variables to link the baseline data to the NDSS, CDR and NHID.

The diagnoses of chronic diseases are based on well-accepted international standards. Causes of death are coded according to the 10th version of the International Statistical Classification of Diseases (ICD-10).<sup>27</sup> In addition, we finished one to two rounds of passive follow-up (varying between sites), and 534 deaths and 3083 new cases of cardiovascular disease (CVD) during 2019–2020 were identified.

Approximately 10% of the participants will be actively followed up every 3 years via face-to-face interviews, to estimate changes in lifestyle and exposure factors and to identify participants who have permanently moved out of the study areas, as well as to minimize the under-reporting of vital status and death data. Biological specimens, including blood, saliva and stool samples, are also planned to be collected. We aim for the rate of loss of individuals to follow-up to be <10%.

## What has been measured?

The types of measures are summarized in Table 1. In brief, a questionnaire survey, physical examination, biological sample collection and laboratory testing were conducted. The entire visit procedure, including the electronic questionnaire interview, medical examination and sample collection, typically took 60–75 min for each participant to complete.

### Questionnaire survey

The face-to-face questionnaire interview collected baseline information about the participants, including demographic and socio-economic characteristics, lifestyle factors, environmental exposures, medical history, mental health status and, for women, reproductive history (Table 1). The questionnaire referred to the China Kadoorie Biobank as a reference<sup>28</sup> and validated instruments for food frequency (Food Frequency Questionnaire) and quality of life [12-item short-form survey (SF-12)] were used.<sup>28–30</sup> A tablet computer with a self-developed application (the RECS App) was used to collect the questionnaire information at face-to-face interviews performed by well-trained technicians. To ensure the quality of the data, interviewers were trained in the standardized process and mode of query to be used.

### Medical examination and sample collection

After the questionnaire survey, 10-mL venous blood samples and anthropometric measurements were taken for all

**Table 1** Information collected in the RECS baseline survey from June 2018 to May 2019

Collected information	No. of variables	Description
Questionnaire survey		
Demographic and socio-economic information	41	Sex, date of birth, ethnic group, education, marital status, occupation, household income, medical insurance coverage, survey date
Lifestyle	225	Tea and coffee drinking, alcohol consumption, smoking history, diet (including Food Frequency Questionnaire), physical activities, sleeping habits
Environmental exposures	37	Passive smoking and indoor air pollution, occupational pollution-exposure history
Medical history	194	Personal and family disease history, self-rated health status
Mental health	45	Life satisfaction, traumatic events, mood status, quality of life (SF-12)
Reproductive history (for women)	47	Age at menarche, menopause status, pregnancy history
Physical examination	16	Height, weight, waist circumference, blood pressure, body composition (BIA)
Biological sample collection	–	Blood samples collected and divided into serum, plasma and buffy coat samples. Buffy coat would be used for the further genetic tests. All samples are stored in RECS Biobank
Laboratory testing	–	Blood, stool and urine tests, blood lipids, blood glucose, liver function, renal function (sub-studies)

BIA, bio-electrical impedance analysis; RECS, Regional Cohort of Northwest China Study; SF-12, 12-item short-form survey.

participants by professional physicians or nurses at the local clinical centre, with participants asked to fast overnight for  $\geq 8$  hours beforehand. The doctors, nurses and technicians received standardized training before conducting the examinations. For all participants, the medical examination included measurement of height, weight, waist circumference (WC), heart rate, body fat percentage, visceral fat index, basal metabolic rate, bone mass, muscle mass, total body water and resting blood pressure. Each project group (the provincial CDC or a local medical college) had primary responsibility for quality control, as guided by the overall protocol of the study. The measurements obtained in the medical examinations were estimated to have optimal consistency among the devices that were used in the five provinces. In addition, blood tests (e.g. liver and kidney function, blood lipids, fasting blood glucose and glycosylated haemoglobin, high-sensitivity C-reactive protein), urine tests, lung-function tests and B-ultrasonic examinations were selectively used in  $\sim 40\%$  of the participants.

## What has it found?

The provinces and autonomous regions in Northwest China largely comprise China's inland poverty belt, with a total land area accounting for 33.3% of the national land area.<sup>13</sup> The population density in the areas suitable for human habitation exceeds the critical level of population density in arid and semi-arid areas (as designated by the UN) and the distribution of gross domestic product (GDP), life expectancy and forest coverage is unbalanced<sup>15</sup> (Table 2).

Table 2 shows the socio-economic characteristics of the participants (in total and in each of the five provinces). Among the total 118 572 participants, the proportions of Han, Uyghur, Kazakh, Hui and Tibetan ethnicities were 75.3%, 13.0%, 1.7%, 8.2% and 1.3%, respectively. Participants in the provinces of Gansu, Shaanxi and Qinghai were primarily of Han ethnicity ( $>90\%$ ), with a small proportion of people of Tibetan ethnicity in the Gansu (7.2%) and Qinghai (1.3%) provinces. In Xinjiang province, 67.2% of participants were of Uyghur, Kazakh or

**Table 2** Socio-economic status of study area and the characteristics of participants in the RECS

Variables	Total	Shaanxi	Xinjiang	Ningxia	Gansu	Qinghai
<b>Socio-economic indicators<sup>a</sup></b>						
Population (million)	101.86	38.35	24.45	6.82	26.26	5.98
Proportion of non-agricultural population (%)	29.8	26.18	40.06	36.49	25.35	28.08
Life expectancy (years)	72.51	74.68	72.32	73.38	72.23	69.96
GDP per capita (Yuan, RMB)	43 762	56 154	46 089	47 177	28 026	41 366
Forest coverage (%)	15.5	43.1	4.9	12.6	11.3	5.8
<b>Characteristics of participants (n)</b>						
Sex [n (%)]						
Male	47 655 (40.2)	18 597 (38.7)	12 572 (40.7)	6378 (40.2)	9003 (43.3)	1105 (36.7)
Female	70 906 (59.8)	29 428 (61.3)	18 280 (59.3)	9500 (59.8)	11 791 (56.7)	1907 (63.3)
Age, years, mean $\pm$ SD						
	52.41 $\pm$ 12.04	49.22 $\pm$ 13.50	55.18 $\pm$ 11.10	56.79 $\pm$ 10.08	52.25 $\pm$ 9.11	52.85 $\pm$ 9.31
Ethnic group [n (%)]						
Han	89 001 (75.3)	47 398 (99.1)	9911 (32.2)	9874 (62.3)	19 009 (92.1)	2809 (93)
Uyghur	15 412 (13.0)	25 (0.1)	15 348 (49.9)	12 (0.1)	24 (0.1)	<10 (0.1)
Kazakh	2031 (1.7)	52 (0.1)	1938 (6.3)	<10 (0)	35 (0.2)	<10 (0)
Hui	9645 (8.2)	184 (0.4)	3376 (11)	5915 (37.3)	38 (0.2)	132 (4.4)
Tibetan	1555 (1.3)	11 (0)	15 (0)	<10 (0)	1483 (7.2)	39 (1.3)
Other ethnic groups <sup>b</sup>	476 (0.4)	163 (0.3)	193 (0.6)	24 (0.2)	59 (0.3)	37 (1.2)
Educational level [n (%)]						
No formal school	21 494 (18.2)	5627 (11.8)	5119 (16.6)	5489 (34.6)	4593 (22.2)	666 (22.1)
Primary school	34 282 (29.0)	7657 (16)	15 459 (50.2)	5199 (32.8)	5366 (25.9)	601 (19.9)
Middle school	42 444 (35.9)	17 720 (37.1)	9038 (29.4)	5069 (32)	9093 (43.9)	1524 (50.5)
College and above	19 884 (16.8)	16 751 (35.1)	1168 (3.8)	96 (0.6)	1642 (7.9)	227 (7.5)
Household annual income <sup>c</sup> [n (%)]						
<¥10 000	21 687 (19.6)	5136 (12.1)	10 387 (34.4)	3898 (25.9)	2131 (10.6)	135 (5)
<¥50 000	60 034 (54.3)	21 340 (50.4)	13 646 (45.2)	9378 (62.3)	14 064 (69.8)	1606 (59)
<¥100 000	19 420 (17.6)	9091 (21.5)	4502 (14.9)	1447 (9.6)	3551 (17.6)	829 (30.4)
$\geq$ ¥100 000	9319 (8.4)	6738 (15.9)	1681 (5.6)	339 (2.3)	408 (2)	153 (5.6)

GDP, gross domestic product; RECS, Regional Cohort of Northwest China Study; RMB, Renminbi; SD, standard deviation.

<sup>a</sup>Data are from the 2018 China Statistical Yearbook, <http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm>.

<sup>b</sup>Other ethnic groups include Mongolian, Man and Tu.

<sup>c</sup>RMB (yuan) was used to estimate household income.

Hui ethnicity, and in Ningxia province, 37.3% of participants were of Hui ethnicity.

Table 3 shows lifestyle and self-reported medical-history factors across the different ethnic groups. Kazakh (30.0%) and Tibetan participants (19.9%) had higher proportions of current smokers. There were also higher proportions of participants who usually drank alcohol in the Han (9.2%) and Tibetan (8.2%) ethnic groups. The participants from the Uyгур, Kazakh, Hui and Tibetan ethnic

groups consumed more animal-based food (but less salt). The overall prevalences of self-reported diabetes and CVD were low in the baseline survey and extremely low among several ethnic groups (other than Han). This may be because there is less access to health services and chronic-disease management.

Based on the Working Group on Obesity in China criteria,<sup>31</sup> the prevalence of overweight [body mass index (BMI), 24 to <28 kg/m<sup>2</sup>] and obesity (BMI ≥28 kg/m<sup>2</sup>)

**Table 3** Lifestyle of participants by ethnic group

Variables	Total (n = 117 644)	Han (n = 89 001)	Uyгур (n = 15 412)	Kazakh (n = 2031)	Hui (n = 9645)	Tibetan (n = 1555)
<b>Sex and age</b>						
Sex [n (%)]						
Male	47 242 (40.2)	35 635 (40)	5919 (38.4)	965 (47.5)	4062 (42.1)	661 (42.5)
Female	70 391 (59.8)	53 356 (60)	9493 (61.6)	1066 (52.5)	5583 (57.9)	893 (57.5)
Age (years) (mean ± SD)	52.43 ± 12.03	52.28 ± 12.43	52.46 ± 10.35	49.28 ± 9.99	54.72 ± 10.93	50.67 ± 11.42
<b>Lifestyle</b>						
Tea consumption [n (%)]						
Never or almost never	57 221 (48.8)	44 530 (50.3)	7263 (47.2)	1072 (52.9)	3535 (36.7)	821 (53.2)
Occasionally	21 546 (18.4)	17 279 (19.5)	2148 (14)	444 (21.9)	1483 (15.4)	192 (12.5)
Usually	38 393 (32.8)	26 781 (30.2)	5961 (38.8)	511 (25.2)	4611 (47.9)	529 (34.3)
Alcohol consumption [n (%)]						
Never or almost never	85 501 (73.1)	59 545 (67.3)	14 094 (92.1)	1577 (77.8)	9095 (94.5)	1190 (77.4)
Occasionally	22 402 (19.1)	20 786 (23.5)	595 (3.9)	399 (19.7)	401 (4.2)	221 (14.4)
Usually	9103 (7.8)	8178 (9.2)	618 (4)	50 (2.5)	131 (1.4)	126 (8.2)
Smoking habits [n (%)]						
Non-smoker	95 096 (81.5)	70 347 (79.8)	13 587 (88.7)	1407 (69.6)	8559 (88.9)	1196 (78.2)
Ceased smoker	3263 (2.8)	2766 (3.1)	225 (1.5)	9 (0.4)	234 (2.4)	29 (1.9)
Current smoker	18 330 (15.7)	15 087 (17.1)	1498 (9.8)	606 (30)	835 (8.7)	304 (19.9)
<b>Diet frequency, times/week</b>						
Animal-based food (mean ± SD)	5.94 ± 5.14	5.49 ± 5.35	7.77 ± 3.46	7.81 ± 3.69	6.60 ± 4.85	7.57 ± 4.62
Plant-based food (mean ± SD)	16.04 ± 6.20	15.67 ± 6.38	17.59 ± 5.51	17.50 ± 4.28	17.06 ± 5.17	13.32 ± 6.07
Eggs and dairy (mean ± SD)	5.32 ± 5.93	5.27 ± 5.77	5.36 ± 6.35	12.90 ± 6.89	4.70 ± 5.35	2.21 ± 5.12
<b>Salt consumption (g/day)</b>						
Household (mean ± SD)	22.79 ± 18.48	20.22 ± 17.99	33.74 ± 18.24	29.42 ± 15.98	23.73 ± 15.65	14.67 ± 19.72
Per capita (mean ± SD)	7.17 ± 6.30	6.62 ± 5.91	9.56 ± 7.96	7.43 ± 4.71	7.86 ± 5.64	3.36 ± 3.69
<b>Physical activity</b>						
Exercise frequency [n (%)]						
Never or almost never	72 163 (62.4)	48 733 (55.9)	14 085 (92.1)	1638 (80.9)	6400 (66.9)	1307 (85.6)
1–3 times/month	10 603 (9.2)	9032 (10.4)	601 (3.9)	266 (13.1)	664 (6.9)	40 (2.6)
1–2 times/week	7187 (6.2)	6786 (7.8)	142 (0.9)	15 (0.7)	214 (2.2)	30 (2)
3–5 times/week	4899 (4.2)	4591 (5.3)	75 (0.5)	7 (0.3)	204 (2.1)	22 (1.4)
Daily or almost every day	20 794 (18.0)	18 085 (20.7)	395 (2.6)	98 (4.8)	2089 (21.8)	127 (8.3)
<b>Sleeping habits</b>						
Sleep duration (hours) (mean ± SD)	7.12 ± 1.34	7.10 ± 1.32	7.16 ± 1.33	7.43 ± 1.27	7.00 ± 1.39	8.07 ± 1.77
<b>Medical history</b>						
Self-reported diabetes [n (%)]						
Yes	5694 (4.8)	4271 (4.8)	739 (4.8)	34 (1.7)	631 (6.5)	19 (1.2)
No	111 950 (95.2)	84 730 (95.2)	14 673 (95.2)	1997 (98.3)	9014 (93.5)	1536 (98.8)
Self-reported CVD [n (%)]						
Yes	25 634 (21.8)	17 279 (19.4)	5442 (35.3)	333 (16.4)	2435 (25.2)	145 (9.3)
No	92 010 (78.2)	71 722 (80.6)	9970 (64.7)	1698 (83.6)	7210 (74.8)	1410 (90.7)

SD, standard deviation; CVD, cardiovascular disease.

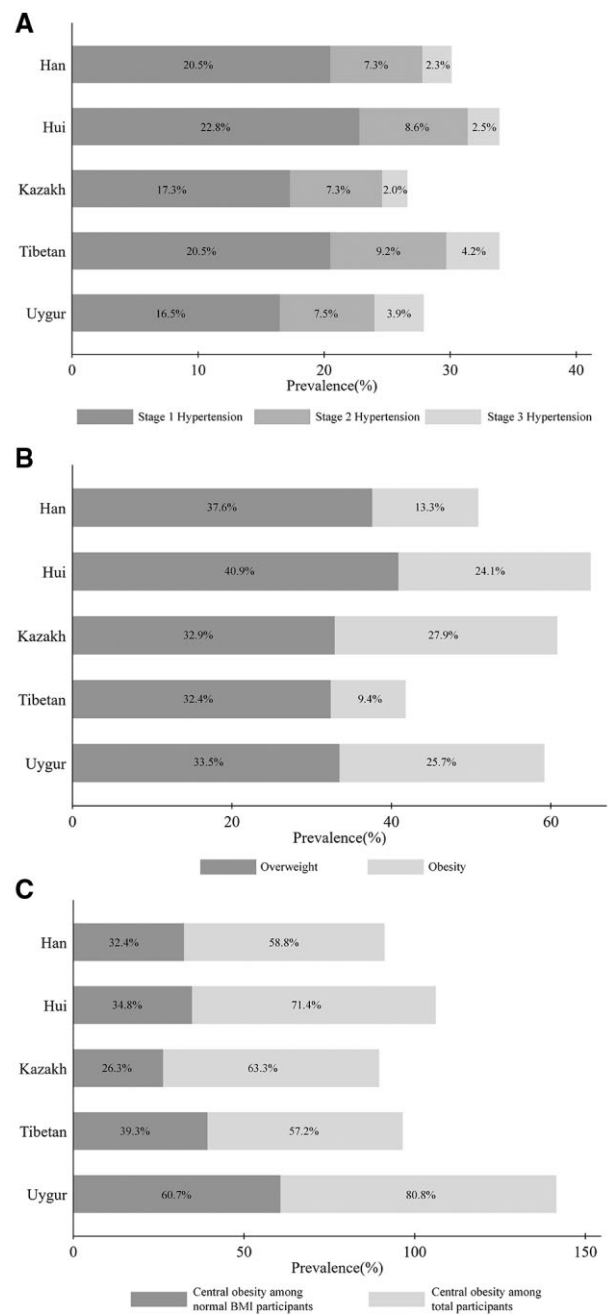
was 37.2% and 16.1%, respectively. Participants from the Uygur, Kazakh and Hui ethnic groups had a higher prevalence of central obesity (WC >85 cm in men or >80 cm in women) compared with those in the Han ethnic group. Even among participants with normal BMI, the prevalence of central obesity, especially in the Uygur ethnic group, was high. The prevalence of hypertension (as defined by the Chinese Guidelines for Prevention and Treatment of Hypertension<sup>32</sup>) was higher in the Hui, Tibetan and Han ethnic groups (>30%) and the total prevalence of hypertension was ~30.2% (Figure 2). These results suggest that overweight/obesity and hypertension have become a prominent public health problem in Northwest China.

Current projects include analysing the predictors of CVD and the cost-effectiveness of CVD preventive health policy by using baseline data as well as a study of 450 plasma samples exploring the metabolic profile and possible metabolic path to normal weight obesity. Additionally, another study is underway to construct a haplotype reference sequence set of different ethnicities of north-western China via high-throughput sequencing based on 1000 ethnic samples of DNA.

### What are the main strengths and weaknesses?

There are several limitations to this cohort. First, information on personal history of diseases, physical conditions and some lifestyle factors (such as smoking, drinking and dietary food intake) was self-reported, and recall bias could not be avoided. Second, we only included participants in the age range of 35–74 years, which may have excluded information on early-life exposures. Third, due to limited funding, only 40% of participants were tested for biochemical indices. Fourth, no information on non-responders was collected, so it is possible that bias might exist in our sample.

As a unique, large-scale cohort in Northwest China, our study also has some strengths. First, because of the specific features regarding lifestyle, dietary habits, climate, geography and rapid social and economic transitions in Northwest China, this cohort will provide important data and support to estimate the prevalence and risk factors for NCDs in Northwest China, which will allow us to further investigate the aetiology of NCDs. Second, this cohort comprises multi-ethnic individuals, urban and rural inhabitants, plateau and basin residents, populations living in remote mountain areas and those living in highly concentrated and air-polluted areas. These features allow us to evaluate the effects of various environmental exposures on different health-related outcomes. Third, many blood samples (~0.88 million), including whole-blood, plasma, serum and buffy coat samples, were collected at baseline. These valuable samples allow us to obtain information on the genetic and epigenetic variants for



**Figure 2** Prevalence of hypertension, obesity, overweight and central obesity among different regional ethnic groups. Central obesity was defined as waist circumference of >85 cm in men or >80 cm in women. Obesity was classified as body mass index (BMI) of >28 kg/m<sup>2</sup>; overweight as BMI of 24.0 to <28 kg/m<sup>2</sup>; and normal weight as BMI of 18.5–23.9 kg/m<sup>2</sup>.

health outcomes. In addition, genetic information from multi-ethnic individuals, combined with abundant personal and environmental information (including specific lifestyle and dietary habits that were collected at baseline), provides an important opportunity to reveal complex and specific mechanisms of the genetic and environmental factors associated with NCDs. Finally, the information on follow-up

outcomes, including the diagnosed diseases that are linked to this cohort database through the NDSS, CDR and NHID systems, enables the reliable and real-time confirmation of participants' health status and the tracking of medical histories.

### Can I obtain the data? Where can I find out more?

The RECS welcomes collaboration throughout the world to maximize the use of these data. Currently, the data cannot be downloaded publicly because they contain sensitive information. However, possible collaborators are invited to contact the corresponding author (Shaonong Dang: [tjdsn@xjtu.edu.cn](mailto:tjdsn@xjtu.edu.cn)).

### Ethics approval

The study was approved by the Human Research Ethics Committee of the Xi'an Jiaotong University Health Science Center (No: XJTU2016-411).

### Funding

This research is supported by National Key Research and Development Program of China [grand nos 2017YFC0907200, 2017YFC0907201, 2017YFC0907202, 2017YFC0907203, 2017YFC0907204 and 2017YFC0907205].

### Acknowledgements

We thank all study participants and staff in each of the study sites and communities, including Xinjiang Provincial Center of Disease Prevention and Control, Traditional Chinese Medical Hospital of Xinjiang Uygur Autonomous Region, Gansu Provincial Center of Disease Prevention and Control, Ershilipu Township Hospital in Qinghai Province, Ningxia Provincial Center of Disease Prevention and Control, Shaanxi Provincial Center of Disease Prevention and Control, The First Affiliated Hospital of Xi'an Jiaotong University, The Second Affiliated Hospital of Xi'an Jiaotong University, Northwest Women's and Children's Hospital, Shaanxi Provincial People's Hospital, Jiren Hospital.

### Author contributions

H.Y. and S.D. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: H.Y., S.D., C.L., L.Z., Q.L. and Y.S. Acquisition, analysis, or interpretation of data: all authors. Drafting of the manuscript: C.L., Y.L. and S.D. Critical revision of the manuscript for important intellectual content: all authors. Statistical analysis: C.L., Y.L., G.S. and S.D. Administrative, technical or material support: G.S., B.M., Y.Zhao, L.P., Y.K., S.X., X.L., J.D., Y.Zhang, X.H.W., X.J.W. and D.W.

### Conflict of interest

We confirm that there is no conflict of interest to declare.

### References

- Niessen LW, Mohan D, Akuoku JK *et al*. Tackling socioeconomic inequalities and non-communicable diseases in low-income and middle-income countries under the Sustainable Development agenda. *Lancet* 2018;391:2036–46.
- World Health Organization. *Preventing Chronic Diseases: A Vital Investment: WHO Global Report*. 2005. [https://www.who.int/chp/chronic\\_disease\\_report/full\\_report.pdf](https://www.who.int/chp/chronic_disease_report/full_report.pdf) (30 August 2021, date last accessed).
- World Health Organization. *Health in 2015: from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals*. 2015. [https://apps.who.int/iris/bitstream/handle/10665/200009/9789241565110\\_eng.pdf?sequence=1&isAllowed=y](https://apps.who.int/iris/bitstream/handle/10665/200009/9789241565110_eng.pdf?sequence=1&isAllowed=y) (30 August 2021, date last accessed).
- United Nations General Assembly. Resolution on the Prevention and Control of Non-Communicable Disease. 2010. [https://www.un.org/ga/search/view\\_doc.asp?symbol=A/64/L.52&Lang=E](https://www.un.org/ga/search/view_doc.asp?symbol=A/64/L.52&Lang=E) (30 August 2021, date last accessed).
- Chen WG, Liu L, Zhu M *et al*. The report of cardiovascular diseases in China—2017 profiles. *Chin Circ J* 2018;33:1–8.
- Pan L, Yang Z, Wu Y *et al*. The prevalence, awareness, treatment and control of dyslipidemia among adults in China. *Atherosclerosis* 2016;248:2–9.
- Xu Y, Wang L, He J *et al*. Prevalence and control of diabetes in Chinese adults. *JAMA* 2013;310:948–59.
- Yan R, Li W, Yin L *et al*. Cardiovascular diseases and risk-factor burden in urban and rural communities in high-, middle-, and low-income regions of china: a large community-based epidemiological study. *J Am Heart Assoc* 2017;6:e004445.
- Center for Health Statistics and Information, National Health and Family Planning Commission. *An Analysis Report of National Health Services Survey in China*. 2013. <http://www.nhc.gov.cn/ewebeditor/uploadfile/2016/10/20161026163512679.pdf> (30 August 2021, date last accessed).
- He X, Wu J, He S. Hydrochemical characteristics and quality evaluation of groundwater in terms of health risks in Luohe aquifer in Wuqi County of the Chinese Loess Plateau, Northwest China. *Hum Ecol Risk Assess Int J* 2019;25:32–51.
- Wu J, Wang L, Wang S *et al*. Spatiotemporal variation of groundwater quality in an arid area experiencing long-term paper wastewater irrigation, Northwest China. *Environ Earth Sci* 2017;76:460.
- Ahmad F, Draz MU, Ozturk I, Su L, Rauf A. Looking for asymmetries and nonlinearities: the nexus between renewable energy and environmental degradation in the northwestern provinces of China. *J Cleaner Prod* 2020;266:121714.
- Cao L, Fang Q, Hong Z *et al*. *Ethnic China: Identity, Assimilation, and Resistance*. Lanham: Lexington Books, 2015.
- Hipgrave DB, Chang S, Li X, Wu Y. Salt and sodium intake in China. *JAMA* 2016;315:703–05.
- Wu A, Zhao J. Non-equilibrium analysis of population distribution in Northwest China from 2001 to 2016. In: Su JJ, Balakrishnan S, Chung MA, eds., *2019 International Conference on Advanced Education, Management and Humanities (AEMH 2019)*. Dordrecht: Atlantis Press, 2019.
- Liu X, Lu L. 2017–20 income and poverty gaps between Han and Ethnic minorities in rural China, 2002 and 2013. In: *Centre for Human Capital and Productivity Working Papers, 2017–20*.



- London, ON: Department of Economics, University of Western Ontario, 2017.
17. Tang L, Shivappa N, Hebert JR, Lee AH, Xu F, Binns CW. Dietary inflammatory index and risk of oesophageal cancer in Xinjiang Uyghur Autonomous Region, China. *Br J Nutr* 2018; **119**:1068–75.
  18. Cai J, Coyte PC, Zhao H. Decomposing the causes of socioeconomic-related health inequality among urban and rural populations in China: a new decomposition approach. *Int J Equity Health* 2017; **16**:128.
  19. Zuo H, Shi Z, Hussain A. Prevalence, trends and risk factors for the diabetes epidemic in China: a systematic review and meta-analysis. *Diabetes Res Clin Pract* 2014; **104**:63–72.
  20. Fei X, Wu J, Kong Z, Christakos G. Urban-rural disparity of breast cancer and socioeconomic risk factors in China. *PLoS One* 2015; **10**:e0117572.
  21. National Bureau of Statistics. *2018 China Statistical Yearbook*. Beijing, China Statistics Press, 2018.
  22. Liu S, Wu X, Lopez AD *et al*. An integrated national mortality surveillance system for death registration and mortality surveillance, China. *Bull World Health Organ* 2016; **94**:46–57.
  23. Yang Y, Sun X, Wang J, Yang C, Zhang L. Incidence rates of four major non-communicable chronic diseases in the Chinese adult population from 2007 to 2016: a study based on a national commercial claims database. *Clin Epidemiol* 2020; **12**: 215–22.
  24. Zheng R, Zeng H, Zhang S, Chen W. Estimates of cancer incidence and mortality in China, 2013. *Cancer Commun* 2017; **36**: 1–6.
  25. Zhang L, Wang H, Li Q, Zhao M, Zhan Q. Big data and medical research in China. *BMJ* 2018; **360**:j5910.
  26. Li Y, Lu C, Liu Y. Medical insurance information systems in China: mixed methods study. *JMIR Med Inform* 2020; **8**: e18780. [PMC][10.2196/18780] [32673209]
  27. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems—10th Revision: Volume 2: Instruction Manual, Fifth edition*. 2016. [https://icd.who.int/browse10/Content/statichtml/ICD10Volume2\\_en\\_2019.pdf](https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf) (30 August 2021, date last accessed).
  28. 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.
  29. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004; **42**:851–59.
  30. Cheng Y, Yan H, Dibley MJ, Shen Y, Li Q, Zeng L. Validity and reproducibility of a semi-quantitative food frequency questionnaire for use among pregnant women in rural China. *Asia Pac J Clin Nutr* 2008; **17**:166–77.
  31. Song P, Li X, Bu Y *et al*. Temporal trends in normal weight central obesity and its associations with cardiometabolic risk among Chinese adults. *Sci Rep* 2019; **9**:5411.
  32. Joint Committee for Guideline Revision. 2018 Chinese Guidelines for Prevention and Treatment of Hypertension: a report of the Revision Committee of Chinese Guidelines for Prevention and Treatment of Hypertension. *J Geriatr Cardiol* 2019; **16**:182–241.