Cohort Profile: The China Northwest Cohort study

Chao Li (lcxjtu@xjtu.edu.cn)
Xi’an Jiaotong University
https://orcid.org/0000-0002-4417-1739

Yezhou Liu
Xi’an Jiaotong University Health Science Center

Guoshuai Shi
Xi’an Jiaotong University Health Science Center

Baibing Mi
Xi’an Jiaotong University Health Science Center

Lingxia Zeng
Xi’an Jiaotong University Health Science Center

Yuxue Bi
Xi’an Jiaotong University Health Science Center

Qiang Li
Xi’an Jiaotong University Health Science Center

Yuan Shen
Xi’an Jiaotong University Health Science Center

Yaling Zhao
Xi’an Jiaotong University Health Science Center

Leilei Pei
Xi’an Jiaotong University Health Science Center

Yijun Kang
Xi’an Jiaotong University Health Science Center

Shengbing Xiao
Xi’an Jiaotong University Health Science Center

Xin Liu
Xi’an Jiaotong University Health Science Center

Jianghong Dai
Xinjiang Medical University

Yuhong Zhang
Ningxia Medical University

Xinhua Wang
Gansu University of Chinese Medicine

Xiaojie Wang
Qinghai Center for Prevention and Control

Duolao Wang
Liverpool School of Tropical Medicine

Hong Yan
Xi’an Jiaotong University Health Science Center

Shaonong Dang
Xi’an Jiaotong University Health Science Center

Study protocol
Abstract

Background

The China Northwest Cohort study (CNC), a community population-based prospective observational study, aimed to investigate specific aetiology causes of NCDs and the long-term health hazards of NCDs in the northwest region of China with different ethnic groups.

Study design and methods

A total of 118,572 participants (aged 35–74 years and including the Han, Uygur, Kazakh, Hui and Tibetan ethnic groups from five provinces of northwest China) 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 with the use of concise questionnaires to review risk exposures and disease incidence, as well as to review biological specimens, including blood, saliva and stool samples, which will be collected. Passive follow-ups will be conducted by periodic linking (every 6 months) of baseline survey data to establish electronic disease registries. The questionnaire survey, a regular medical examination and the storage of blood samples were conducted in the CNC baseline survey for all of the participants. Several other items from the medical examination were recorded for approximately 40% of the participants.

Discussion

Among the total participants, the proportions of Han, Uygur, Kazakh, Hui and Tibetan ethnicities were 75.3%, 13.0%, 1.7%, 8.2% and 1.3%, respectively. Many lifestyle and medical history factors were different across the ethnic groups. The genetic information from the multi-ethnic individuals, combined with abundant personal and environmental information, provided an important opportunity to reveal complex and specific mechanisms of the genetic and environmental factors associated with NCDs.

Background

Why was the cohort set up?

Non-communicable diseases (NCDs) contribute to more than two-thirds of worldwide deaths. Approximately 80% of NCDs occur in low-income and middle-income countries, and 33% of the deaths from NCDs affect people younger than 60 years(1–4). Although great efforts have been made to control NCDs in China, the prevalence of NCDs still continues to increase(5–8). Previous studies have reported that the prevalence of NCDs has increased from 15.3% in 2003 to 29.5% in 2013 in China(9). The burden of NCDs and the associated healthcare costs are significantly rising for the elderly individuals with chronic diseases who require long-term medical services and support. NCD patients are facing risks 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(5, 9).

Northwest China is located in the innermost centre (31°-50°N, 73°-111°E), and this region is considered to be relatively more poor (in financial aspects) than the central and eastern provinces, including Shaanxi, Gansu, Qinghai, Ningxia Hui Autonomous Region and Xinjiang Uygur Autonomous Region(10). The topography is highly varied and includes large stretches of arid desert, wastelands, fertile oases, grassy plateaus and high mountain ranges. Most provinces and autonomous regions in the northwest region of China comprise China's inland poverty belt. This area has jurisdiction over 360 county-level administrative units, with a total land area accounting for 33.27% of the national land area(11). The population density of the suitable areas for human habitation exceed the critical level of population density in arid and semi-arid areas (as designated by the United Nations), and the distribution of the population is unbalanced (Shaanxi: 38.4 million people; Gansu: 25.3 million people; Xinjiang: 24.4 million people; Ningxia: 6.82 million people; and Qinghai: 5.98 million people) (12). These provinces also face various environmental issues, due to decreased water availability and an arid climate(13, 14). In addition, Northwest China is also one of the most ethnically diverse areas in China. Lifestyle and dietary habits differ according to ethnic groups, and the high rates of smoking and drinking in many ethnic populations further increase the risk of NCDs(11, 12, 15–17).
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\(^\text{(18–20)}\). To address the urgent need for data on the prevalence, risk factors and associated conditions of NCDs in resource-constrained settings, the China Northwest Cohort study (CNC), which was supported by the National Key R&D Program of China (NKRP), aimed to establish a nationally representative cohort study including different ethnicities and regions. The China Northwest Cohort study, as one important part of the NKRP, will provide a reliable source of data for investigating specific aetiology causes of NCDs and the long-term health hazards of NCDs in the northwest region of China with different ethnic groups. The results from the cohort study 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.

**Study Design And Methods**

**Who is in the cohort?**

This study was conducted in 13 regions of 5 western provinces (Xinjiang: Moyu county, Yining city, Urumchi city; Gansu: Wuwei city, Tianshui city, Gannan city, Baiying city; Qinghai: Xining city; Ningxia: Qingtongxia city, Pinglou city; Shaanxi: Zhenba county, Baoji city, Xi'an city.) of Northwest China, with the regions chosen according to local disease patterns, exposures to certain risk factors, the distribution of ethnic groups, population size and stability, quality of death and disease registries and local commitment and capacity (Figure 1). This study was designed to include 100,000 participants and was 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), with each group responsible for the baseline survey measurements, including the questionnaire interview, the blood sample collection and storage and further follow-up work. The Shaanxi project team worked as the general manager in charge of the operation of the project. The baseline survey was conducted between June 2018 and May 2019.

A baseline survey was conducted via face-to-face interviews and clinic visits, with the assistance from local health staffs. The samples in the cohort consisted of permanent residents (aged 35-74 years), and any individuals attending the baseline survey who were slightly outside of the target age range were not turned away from the study, in order to encourage participation. In addition, we excluded the participants who had severe mental diseases or disabilities that hindered their ability to communicate in order to obtain accurate data. All the participants were asked to bring their national identity cards in order to exclude individuals who were not local permanent residents because these individuals were unavailable to be followed. All of the participants signed the informed written consents, which allowed us to access their medical records and their long-term storage supplies of blood samples for solely medical research purposes.

**How often will participants be followed?**

The National Death Surveillance System (NDSS), Chronic Disease Registries (CDR), The National Central Cancer Registry (NCCR) and National Health Insurance Claim Datasets (NHICD) enabled us to conduct passive follow-ups by periodic linking (every 6 months) of the baseline survey data to the NDSS, CDR, NCCR and NHICD. The NDSS and CDR are both based on the China’s Disease surveillance points system. The causes of death of the study participants were regularly monitored through the use of official death certificates that were reported to the regional Center for Disease Control and Prevention (CDC), wherein the NDSS was based; additionally, the NDSS was available in all 13 study regions. Information on chronic diseases (diabetes, hypertension, heart failure, stroke and cancer, among other diseases) is currently being collected through linkage with the NCCR, CDR and NHICD. The new national health insurance claim datasets are now fully established in all 13 study regions, and future incidences of chronic disease and the hospitalization information will be primarily extracted from the NHICD. Because the quality of the data from the NCCR and CDR is insufficient, the NCCR and CDR were not primary data sources but were complementary sources for the 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 NHICD. The diagnoses of these diseases are based on well-accepted international standards. The causes of death are coded according to the 10\(^{\text{th}}\) version of the international statistical classification of diseases (ICD-10)\(^{\text{(21)}}\). Approximately 10% of the participants will be actively followed every 3 years via face-to-face interviews, in order to estimate the
changes of lifestyle and exposure factors and to identify the participants who have permanently moved out of the study regions, as well as to minimize the under-reporting of vital status and death data. In addition, biological specimens, including blood, saliva and stool samples, are planned to be collected. The rate of loss of individuals to the follow-up will aim to be less than 10%.

**What has been measured?**

In general, all the participants were asked to fast overnight for at least 8 hours before arriving at the local clinical centre, after which questionnaire surveys, physical examinations, laboratory tests and biological sample collections were conducted. A tablet computer with a self-developed application (the CNC App) was used to collect the questionnaire information, of which the main content referred to the study of the China Kadoorie Biobank as a reference(22), with the information being collected via face-to-face interviews that were performed by well-trained technicians. In addition, blood samples and anthropometric measurements were performed by professional physicians or nurses. The entire visiting procedure, including the electronic questionnaire interview, medical examinations and sample collections, typically took 60-75 min for each participant to complete.

**Questionnaire survey**

The face-to-face questionnaire interview was conducted to collect baseline information about the participants, including demographic and social-economic characteristics (sex, date of birth, ethnic group, education, marital status, occupation and income, among other factors), lifestyle factors (tea and coffee drinking habits, alcohol consumption, smoking history and dietary status, among other factors), environmental exposures (passive smoking and indoor air pollution, among other factors), medical history, mental health status and reproductive history (Table 1). To ensure the quality of the data from the questionnaire survey, reviewers were trained on the unifying process and the mode of query.

**Medical examinations**

The medical examinations were primarily conducted by using the resources and personnel at the local clinical centres. Standardized training for the doctors, nurses and technicians was implemented before the investigation. The items of the medical examinations for all of the participants included height, weight, waist circumference, heart rate, body fat percentage (BFP), visceral fat index (VFI), basal metabolic rate (BMR), bone mass, muscle mass, total body water (TBW) and resting blood pressure. The project group (the provincial CDC or a local medical college) had the primary responsibility of quality control, as guided by the alternative methods of the study. The values of the medical examinations were estimated to have optimal consistency among the devices that were used in 5 provinces. In addition, other items (lung function, blood and urine routine examinations, liver and kidney function tests, glycosylated haemoglobin measurements and B-ultrasonic examinations, among other tests) have been selectively tested in a portion of the participants (approximately 40%).

**Blood samples**

After at least 8 hours of fasting, 10 ml venous blood samples were collected from all of the participants at the time of the baseline survey. The venous blood was transferred into one 5.0 mL ethylenediamine tetra acetic acid dipotassium (EDTA K2) anticoagulation tube and one 5.0 mL vacuum tube without anticoagulation. Plasma, serum and buffy coat samples were separated from the whole blood via centrifugation for 10 min at the relative centrifugal force of 3000 g at 4 °C. Due to limited funds, partial serum samples (approximately 40%) were forwarded to measure various biochemical indexes, including fasting blood glucose, blood lipids, kidney function and high-sensitivity C-reactive protein (hsCRP) tests, among other indexes, within 2 hours.

During the baseline survey period, five provincial biobanks were established, which were responsible for the storage and management of the biological samples of the CNC. All of the blood samples were stored in a -20 °C freezer before being transferred to be stored at -80 °C in cryogenic refrigerators that were located in each provincial biobank in the local CDC, medical college or affiliated hospital of the university via cold chain. To ensure the safety of the biological samples, a portion of the blood samples (including 5 ml plasma and 5 ml whole blood of the participants that were stored in four other provincial Biobanks [except for the Xi’an biobank]) were required to be transferred to the Xi’an biobanks in Shaanxi province as a back-up storage. Due to the long distance between the Shaanxi province and the other 4 provincial sites (especially Xinjiang), the quality control of each
shipment was a key issue that had to be specifically considered. To prevent any blood sample from thawing during the shipment from each site to the Xi’an biobank, an electronic thermometer was placed in each sample’s dry ice box, in order to dynamically monitor the temperature; additionally, the temperature of each of the samples was maintained at a range of -60–80 °C in the old chain trucks.

A computer-based scanning and packing system for the biological samples was developed, in order to assist in blood sample entry into the cryogenic refrigerators in the biobank. Moreover, quality control requirements were strictly followed throughout the process of blood sample collection, shipment and storage, such as checking the identity of the sample and the participant ID, regulating the storage time and standardizing the volume of the samples and the temperature conditions of the samples on site, in order to ensure high-quality blood samples. In total, approximately 0.88 million blood samples (including plasma, serum, buffy coat and whole blood samples) have been stored in the CNC biobank.

### Discussion

**What has it found?**

Table 2 shows the results of the social-economic characteristics of the participants (in total and across the five provinces). In general, almost 120,000 participants were involved in this cohort in Northwest China. Among the total participants, the proportions of Han, Uygur, Kazakh, Hui and Tibetan ethnicities were 75.3%, 13.0%, 1.7%, 8.2% and 1.3%, respectively. There were more female participants (59.8%) than male participants, the mean age of the participants was 52.41 years and the majority of the participants were of Han ethnicity, had middle school educational levels and were employed. Nearly 75% of the participants had a household income less than 50,000 yuan (RMB) per year. In addition, the ethnicities of the investigated participants in the provinces of Gansu, Shaanxi and Qinghai were of the Han ethnicity (over 90%), with the Xinjiang and Ningxia provinces having 67.2% (Uygur, Kazakh and Hui) and 37.3% (Hui) participants of the muslin ethnicity and a small proportion of the Tibetan ethnicities being observed in the Gansu (7.2%) and Qinghai (1.3%) provinces. The Shaanxi province was the main contributor to the CNC cohort, due to it having the largest population. The participants in the Shaanxi province had better education levels and higher household income than the other four provinces. The proportions of the participants with college or above education levels and those participants with a household income over 100,000 yuan (RMB) per year were 35.1% and 15.9%, respectively, in the Shaanxi province. These proportions among the other four provinces ranged from 0.6-7.9% for college and above education and from 2.0-5.6% for a household income over 100,000 yuan (RMB).

Table 3 reports the lifestyle and medical history factors across the different ethnic groups. Specifically, the participants from the Hui and Uygur ethnic groups preferred to drink tea, with the proportion of individuals usually drinking tea being 47.9% and 38.8%, respectively. Kazakh (30.0%) and Tibetan participants (19.9%) had a higher proportion of current smokers. There was also a higher proportion of participants who usually drank alcohol in the Han (9.2%) and Tibetan (8.2%) ethnic groups. The participants from the Uygur, Kazakh, Hui and Tibetan ethnic groups consumed more animal food (but less salt), compared with the Han ethnic group; however, the frequency of plant food intake among Tibetan participants was the lowest (compared to the other groups), and egg and dairy foods were most frequently eaten by participants in the Kazakh ethnic group. The proportion of daily exercise was higher in the Hui (21.8%) and Han (20.7%) ethnic groups. In addition, the overall prevalences of self-reported diabetes and CVD were low in the baseline survey, and these prevalences were extremely low among several ethnic groups (other than Han ethnicity). This may be related to the fact that there is less access to health service and chronic disease management; thus, the improvement of healthcare utilization should be an urgent consideration in Northwestern China. For life satisfaction, the results showed that more participants of other ethnicities were very satisfied with their life than those participants of the Han ethnicity. The number of pregnancy and delivery cases was 2.91 and 2.42 times, respectively, in the number of total participants. When compared with the Han ethnic group, the women in the other four ethnic groups had more incidences of pregnancy and delivery. In addition, the prevalences of being overweight, as well as obesity and hypertension, were 37.2%, 16.1% and 30.1%, respectively, among the total number of participants. The participants from the Uygur, Kazakh, Hui and Tibetan ethnic groups had higher prevalences of being overweight (24≤BMI<28) and being obese (BMI ≥28), when compared with those in the Han ethnic group. Similar to the other factors of lifestyle and medical history, the prevalence of hypertension (as defined by the Chinese guidelines for prevention and treatment of hypertension[23]) was also different across the five ethnic groups. In general, the prevalence of hypertension was higher in the Hui, Tibetan and Han ethnic groups (> 30%), and there was a higher proportion of stage 2 and
stage 3 hypertension in these ethnic groups. The prevalences of stage 2 and stage 3 hypertension were 13.4%, 11.4% and 11.1% in the Tibetan, Uygur and Hui ethnic groups, respectively.

What are the main strength and weakness?

There are several limitations and strengths that need to be addressed. First, the information on the personal history of diseases, physical conditions and some lifestyle factors (such as smoking, drinking and dietary food intake, among other factors) were self-reported, and recall bias could not be avoided. Second, we only included the participants who were aged in the range of 35 to 74 years, which may have ignored the information on early life exposures. Third, only 40% of the participants were measured for biochemical indexes, due to limited funding. To the best of our knowledge, as being a unique, large-scale cohort study conducted in Northwest China, our study 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 study will provide important data and support in estimating the prevalence and risk factors of NCDs in Northwest China, which will allow us to further provide an opportunity of broadly investigating the aetiology of NCDs. Second, this cohort was comprised of multi-ethnic individuals, urban and rural inhabitants, plateau and basin residents, populations living in remote mountains areas and individuals living in highly concentrated and air polluted areas. These features allowed us to evaluate the effects of various environmental exposures on different health-related outcomes. Third, the large number of obtained blood samples (approximately 0.88 million), including whole blood, plasma, serum and buffy coat samples, were collected at the baseline measurements. These valuable samples allowed us to obtain and identify information regarding the genetic and epigenetic variants for health outcomes. In addition, the genetic information from the multi-ethnic individuals, combined with abundant personal and environmental information (including specific lifestyle and dietary habits that were collected in the baseline measurements), provided an important opportunity to reveal complex and specific mechanisms of the genetic and environmental factors associated with NCDs. Finally, the information on the follow-up outcomes, including the diagnosed disease that are linked to this cohort database through the NDSS, CDR and NHICD systems, enabled the reliable and real-time confirmation of the participants’ health statuses and the tracking of medical histories.

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

The China Northwest Cohort Study welcomes collaboration throughout the world to maximize the use of these data. Currently, the data are not available because they contain some sensitive information. However, possible collaborators are invited to contact the corresponding author (Shaonong Dang: tjdshn@xjtu.edu.cn).

Profile in a nutshell

- The China Northwest Cohort study (CNC), a community population-based prospective observational study, aimed to investigate specific aetiology causes of NCDs and the long-term health hazards of NCDs in the northwest region of China with different ethnic groups.

- A total of 118,572 participants (aged 35-74 years and including the Han, Uygur, Kazakh, Hui and Tibetan ethnic groups from five provinces of northwest China) 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 with the use of concise questionnaires to review risk exposures and disease incidence, as well as to review biological specimens, including blood, saliva and stool samples, which will be collected. Passive follow-ups will be conducted by periodic linking (every 6 months) of baseline survey data to establish electronic disease registries.

- The questionnaire survey, a regular medical examination and the storage of blood samples were conducted in the CNC baseline survey for all of the participants. Several other items from the medical examination were recorded for approximately 40% of the participants.

- Collaborations from all over the world are welcome. Please contact the corresponding author with specific research ideas or questions at: tjdshn@xjtu.edu.cn.

Declarations

Ethics approval and consent to participate
The study was conducted in accordance with the declaration of Helsinki and approved by the Human Research Ethics Committee of the Xi’an Jiaotong University Health Science Center (No: XJTU2016-411). Written informed consent was obtained from individual or guardian participants.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Competing interests

The authors declare no competing interests.

Funding

This research is supported by National Key R&D Program of China (Grand No: 2017YFC0907200, 2017YFC0907201, 2017YFC0907202, 2017YFC0907203, 2017YFC0907204 and 2017YFC0907205)

Authors’ contributions

CL and YL did the analysis and drafted the initial manuscript. GS, BM, LZ, QL. YS, YZ, LP, YK, SX, XL, JD, YZ, XW critically reviewed and revised the manuscript. DW critically reviewed the results of analyses, and reviewed and revised the manuscript. All authors were responsible for the study concept and design and approved the final draft of the manuscript. ND and HY is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

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.

References

1. 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.
2. Organization WH. Preventing chronic diseases: a vital investment: WHO global report. 2005.
3. Organization WH. Health in 2015: from MDGs, millennium development goals to SDGs, sustainable development goals. 2015.
4. UN. Prevention and control of non-communicable disease. 2010.
5. Chen W GR, Liu L, Zhu M, et al. The report of cardiovascular diseases in China–2017 profiles. Chinese Circulation Journal 2018;33:1-8.
6. 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.
7. Xu Y, Wang L, He J, et al. Prevalence and control of diabetes in Chinese adults. JAMA 2013;310:948-59.
8. 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.
9. (NHFPC) CfHSaI. An Analysis Report of National Health Services Survey in China. 2013.
10. 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. Human and Ecological Risk Assessment: An International Journal 2019;25:32-51.
11. Cao L, Fang Q, Hong Z, et al. Ethnic China: Identity, Assimilation, and Resistance. Lexington Books 2015.
12. Wu A and Zhao J. Non-equilibrium Analysis of Population Distribution in Northwest China from 2001 to 2016. 2019 International Conference on Advanced Education, Management and Humanities (AEMH 2019) 2019: Atlantis Press.
13. 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. Environmental Earth Sciences 2017;76:460.
14. Ahmad F, Draz MU, Ozturk I, et al. Looking for asymmetries and nonlinearities: The nexus between renewable energy and environmental degradation in the Northwestern provinces of China. Journal of Cleaner Production 2020;266:121714.
15. Hipgrave DB, Chang S, Li X, et al. Salt and Sodium Intake in China. JAMA 2016;315:703-5.
16. Liu X, Lu L. 2017-20 Income and Poverty Gaps between Han and Ethnic Minorities in Rural China, 2002 and 2013. CHCP Working Paper 2017.
17. Tang L, Shivappa N, Hebert JR, et al. Dietary inflammatory index and risk of oesophageal cancer in Xinjiang Uyghur Autonomous Region, China. British Journal of Nutrition 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. International journal for equity in 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, et al. Urban-rural disparity of breast cancer and socioeconomic risk factors in China. PLoS One 2015;10:e0117572.
21. Organization WH. International statistical classification of diseases and related health problems: instruction manual. 2004.
22. 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.

Tables

Table1 Information collected in CNC baseline survey
| Collected information                        | No. of variables | Description                                                                                                                                 |
|---------------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| **Questionnaire survey**                    |                  |                                                                                                                                             |
| Demographic and social economic information| 41               | Sex, date of birth, ethnic group, education, marital status, occupation, household income, medical insurance coverage, survey date, etc.    |
| Life style                                  | 225              | Tea & coffee drinking, Alcohol consumption, Smoking history, Diet (including food frequency questionnaire), Physical activities, sleeping habits, etc. |
| Environmental exposure                      | 37               | Passive smoking and indoor air pollution, occupational polluters exposure history, etc.                                                     |
| Medical history                             | 194              | Personal and family disease history, self-rated health status, defecation and gum situation, etc.                                             |
| Mental health                               | 45               | Life satisfaction, traumatic events, mood status, life quality(SF-12), etc.                                                                 |
| Reproductive history (for women)            | 47               | Age of menarche, menopause status, pregnancy history, etc.                                                                                   |
| **Physical examination**                    | 16               | Height, weight, waist circumference, blood pressure, body composition (BIA), etc.                                                           |
| Biological sample collection                | /                | Blood samples had been collected and divided into serum, plasma and buffy coat samples. All samples are stored in CNC Biobank.                   |
| Laboratory test                             | /                | Blood & stool & urine regular test, blood lipids, blood glucose, liver function, etc. (sub-studies)                                            |

Abbreviations: CNC, Cohort of northwest China, SF-12, 12-item short form survey.

**Table 2 Social-demographic characteristics of participants among five provinces/districts**
| Variables                              | Total (n=118572) | Shaanxi (n=48025) | Xinjiang (n=30852) | Ningxia (n=15878) | Gansu (n=20794) | Qinghai (n=3023) |
|----------------------------------------|------------------|-------------------|--------------------|-------------------|-----------------|-----------------|
| **Sex, n(%)**                          |                  |                   |                    |                   |                 |                 |
| Male                                   | 47655 (40.2)     | 18597 (38.7)      | 12572 (40.7)       | 6378 (40.2)       | 9003 (43.3)     | 1105 (36.7)     |
| Female                                 | 70906 (59.8)     | 29428 (61.3)      | 18280 (59.3)       | 9500 (59.8)       | 11791 (56.7)    | 1907 (63.3)     |
| **Age, years, Mean±SD**                | 52.41±12.04      | 49.22±13.50       | 55.18±11.10        | 56.79±10.08       | 52.25±9.11      | 52.85±9.31      |
| **Ethnic group, n(%)**                 |                  |                   |                    |                   |                 |                 |
| Han                                    | 89001 (75.3)     | 47398 (99.1)      | 9911 (32.2)        | 9874 (62.3)       | 19009 (92.1)    | 2809 (93)       |
| Uygur                                  | 15412 (13.0)     | 25 (0.1)          | 15348 (49.9)       | 12 (0.1)          | 24 (0.1)        | 3 (0.1)         |
| Kazakh                                 | 2031 (1.7)       | 52 (0.1)          | 1938 (6.3)         | 6 (0)             | 35 (0.2)        | 0 (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)            | 1538 (5)           | 6 (0)             | 35 (0.2)        | 0 (0)           |
| Other ethnic groups 1                  | 476 (0.4)        | 163 (0.3)         | 193 (0.6)          | 24 (0.2)          | 59 (0.3)        | 37 (1.2)        |
| **Educational level, n(%)**            |                  |                   |                    |                   |                 |                 |
| No formal school                       | 21494 (18.2)     | 5627 (11.8)       | 5119 (16.6)        | 5489 (34.6)       | 4593 (22.2)     | 666 (22.1)      |
| Primary school                         | 34282 (29.0)     | 7657 (16)         | 15459 (50.2)       | 5199 (32.8)       | 5366 (25.9)     | 601 (19.9)      |
| Middle school                          | 42444 (35.9)     | 17720 (37.1)      | 9038 (29.4)        | 5069 (32)         | 9093 (43.9)     | 1524 (50.5)     |
| College and above                      | 19884 (16.8)     | 16751 (35.1)      | 1168 (3.8)         | 96 (0.6)          | 1642 (7.9)      | 227 (7.5)       |
| **Family size, n, Mean±SD**            | 3.65±2.11        | 3.71±2.16         | 3.69±1.89          | 2.87±1.50         | 3.98±2.56       | 4.13±1.83       |
| **Marital status, n(%)**               |                  |                   |                    |                   |                 |                 |
| Married                                | 104632 (88.8)    | 41840 (87.8)      | 26224 (85.4)       | 14717 (93)        | 19107 (92.6)    | 2744 (91.1)     |
| Widowed                                | 7223 (6.1)       | 2082 (4.4)        | 2968 (9.7)         | 980 (6.2)         | 999 (4.8)       | 194 (4.4)       |
| Separated/Divorced                     | 2393 (2.0)       | 526 (1.1)         | 1377 (4.5)         | 69 (0.4)          | 365 (1.8)       | 56 (1.9)        |
| Never married                          | 3574 (3.0)       | 3190 (6.7)        | 135 (0.4)          | 60 (0.4)          | 170 (0.8)       | 19 (0.6)        |
| **Occupation, n(%)**                   |                  |                   |                    |                   |                 |                 |
| Unemployed                             | 30628 (26.1)     | 18245 (38.7)      | 5056 (16.5)        | 1756 (11.1)       | 5012 (24.4)     | 559 (18.5)      |
| Employed                               | 80021 (68.3)     | 26362 (56)        | 22734 (74.2)       | 13719 (86.6)      | 15203 (74)      | 2003 (66.4)     |
| Other or not stated                     | 6498 (5.5)       | 2492 (5.3)        | 2864 (9.3)         | 366 (2.3)         | 320 (1.6)       | 456 (15.1)      |
| **Household annual income, n (%)**     |                  |                   |                    |                   |                 |                 |
| < ¥10,000                              | 21687 (19.6)     | 5136 (12.1)       | 10387 (34.4)       | 3898 (25.9)       | 2131 (10.6)     | 135 (5)         |
| < ¥50,000                              | 60034 (54.3)     | 21340 (50.4)      | 13646 (45.2)       | 9378 (62.3)       | 14064 (69.8)    | 1606 (59)       |
| < ¥100,000                             | 19420 (17.6)     | 9091 (21.5)       | 4502 (14.9)        | 1447 (9.6)        | 3551 (17.6)     | 829 (30.4)      |
| ≥ ¥100,000                             | 9319 (8.4)       | 6738 (15.9)       | 1681 (5.6)         | 339 (2.3)         | 408 (2)         | 153 (5.6)       |

Abbreviations: SD, standard deviation.

1 Other ethnic groups including Mongolian, Man, Tu, etc.
RMB (yuan) was used to estimate household income.

Table 3 Life style and medical history of participants by ethnic group
| Variables                  | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| Sex, n(%)                 |                  |               |                 |                 |              |                 |
| Male                       | 47242(40.2)      | 35635(40)     | 5919(38.4)      | 965(47.5)       | 4062(42.1)   | 661(42.5)       |
| Female                     | 70391(59.8)      | 53356(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     |

**Life style**

**Tea/Alcohol/Smoking using**

**Tea consumption, n(%)**

|                | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| Never or almost never | 57221(48.8)    | 44530(50.3)   | 7263(47.2)      | 1072(52.9)      | 3535(36.7)   | 821(53.2)       |
| Occasionally    | 21546(18.4)      | 17279(19.5)   | 2148(14)        | 444(21.9)       | 1483(15.4)   | 192(12.5)       |
| Usually         | 38393(32.8)      | 26781(30.2)   | 5961(38.8)      | 511(25.2)       | 4611(47.9)   | 529(34.3)       |

**Alcohol consumption, n(%)**

|                | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| Never or almost never | 85501(73.1)    | 59545(67.3)   | 13087(86.7)     | 1407(69.6)      | 8599(89.9)   | 1196(78.2)      |
| Occasionally    | 22402(19.1)      | 20786(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(%)**

|                | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| Non-smoker     | 95096(81.5)      | 70347(79.8)   | 13587(88.7)     | 1407(69.6)      | 8599(89.9)   | 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 | 18330(15.7)      | 15087(17.1)   | 1498(9.8)       | 606(30)         | 835(8.7)     | 304(19.9)       |

**Diet frequency, times/week**

|                | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| Animal 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 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      |
| Egg & dairy food, 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       |

**Salt consumption, g/day**

|                | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| 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       |

**Physical activity**

**Exercise frequency, n(%)**

|                | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|----------------|------------------|---------------|-----------------|-----------------|--------------|-----------------|
| Never or almost never | 72163(62.4)   | 48733(55.9)   | 14085(92.1)     | 1638(80.9)      | 6400(66.9)   | 1307(85.6)      |
| 1-3 times/month  | 10603(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)           |
### Variables

| Variables                             | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|---------------------------------------|------------------|--------------|-----------------|-----------------|--------------|-----------------|
| 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 everyday              | 20794(18.0)      | 18085(20.7)  | 395(2.6)        | 98(4.8)         | 2089(21.8)   | 127(8.3)         |

### Sleeping habits

| 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       |

### Medical history

#### Self-reported diseases history

| Diabetes, n(%)                        |                  |              |                 |                 |              |                 |
|---------------------------------------|------------------|--------------|-----------------|-----------------|--------------|-----------------|
| Yes                                   | 5694(4.8)        | 4271(4.8)    | 739(4.8)        | 34(1.7)         | 631(6.5)     | 19(1.2)         |
| No                                    | 111950(95.2)     | 84730(95.2)  | 14673(95.2)     | 1997(98.3)      | 9014(93.5)   | 1536(98.8)      |

| CVD, n(%)                             |                  |              |                 |                 |              |                 |
|---------------------------------------|------------------|--------------|-----------------|-----------------|--------------|-----------------|
| Yes                                   | 25634(21.8)      | 17279(19.4)  | 5442(35.3)      | 333(16.4)       | 2435(25.2)   | 145(9.3)        |
| No                                    | 92010(78.2)      | 71722(80.6)  | 9970(64.7)      | 1698(83.6)      | 7210(74.8)   | 1410(90.7)      |

### Mental health

#### Satisfaction with life, n(%)

| Very satisfied                        | 31166(27.0)      | 13451(15.4)  | 13262(86.9)     | 1369(67.9)      | 2642(27.6)   | 442(29.2)       |
| Satisfied                             | 62408(54.0)      | 54676(62.7)  | 1583(10.4)      | 502(24.9)       | 4914(51.4)   | 733(48.4)       |
| Neither satisfied nor dissatisfied    | 19450(16.8)      | 16899(19.4)  | 337(2.2)        | 142(7)          | 1776(18.6)   | 296(19.5)       |
| Unsatisfied                           | 2174(1.9)        | 1902(2.2)    | 15(0.1)         | 3(0.1)          | 212(2.2)     | 42(2.8)         |
| Very unsatisfied                      | 352(0.3)         | 274(0.3)     | 57(0.4)         | 1(0)            | 18(0.2)      | 2(0.1)          |

### Reproductive history (for women)

| Numbers of pregnancy, Mean±SD        | 2.91±1.96        | 2.52±1.60    | 4.84±2.72       | 3.25±1.74       | 3.40±1.62    | 2.54±1.02       |
| Numbers of delivery, Mean±SD         | 2.42±1.48        | 2.12±1.20    | 3.80±2.03       | 2.86±1.49       | 2.74±1.31    | 2.33±0.85       |

### Physical examination

| BMI, kg/m², Mean±SD                   | 24.54±3.70       | 24.27±3.47   | 25.40±4.49      | 25.70±4.44      | 25.55±3.71   | 23.55±3.29      |

| BMI level, n(%)                       |                  |              |                 |                 |              |                 |
| <18.5 kg/ m²                          | 3615(3.2)        | 2756(3.2)    | 571(3.7)        | 46(2.3)         | 175(1.9)     | 67(4.3)         |
| 18.5-24 kg/ m²                        | 49860(43.6)      | 39526(45.8)  | 5650(37.1)      | 743(36.9)       | 3110(33.2)   | 831(53.8)       |
| 24-28 kg/m²                           | 42572(37.2)      | 32474(37.6)  | 5103(33.5)      | 661(32.9)       | 3833(40.9)   | 501(32.4)       |
| ≥28 kg/m²                             | 18386(16.1)      | 11508(13.3)  | 3916(25.7)      | 561(27.9)       | 2255(24.1)   | 146(9.4)        |
| Variables                                      | Total (n=117644) | Han (n=89001) | Uygur (n=15412) | Kazakh (n=2031) | Hui (n=9645) | Tibetan (n=1555) |
|-----------------------------------------------|------------------|---------------|-----------------|-----------------|-------------|-----------------|
| SBP, mmHg, Mean±SD                            | 127.69±19.79     | 127.44±19.13  | 127.35±23.40    | 127.02±20.13    | 130.53±18.76 | 128.54±21.48    |
| DBP, mmHg, Mean±SD                            | 80.00±11.84      | 80.49±11.50   | 76.51±12.95     | 78.43±11.86     | 81.21±11.79  | 81.59±12.82     |

**Blood pressure level**, n(%)  

| Category                      | Total   | Han     | Uygur   | Kazakh  | Hui     | Tibetan |
|-------------------------------|---------|---------|---------|---------|---------|---------|
| **Normal**                    | 80084(69.9) | 60388(69.8) | 11020(72.1) | 1465(73.4) | 6196(66) | 1015(66.1) |
| **Optimal**                   | 36007(31.4) | 26329(30.5) | 6179(40.5)  | 734(36.8)  | 2346(25) | 419(27.3)  |
| **Prehypertension**           | 44077(38.4) | 34059(39.4) | 4841(31.7)  | 731(36.6)  | 3850(41) | 596(38.8)  |
| **Hypertension**              | 34563(30.1) | 26069(30.2) | 4255(27.9)  | 531(26.6)  | 3188(34) | 520(33.9)  |
| **Stage 1 Hypertension**      | 23024(20.1) | 17702(20.5) | 2519(16.5)  | 346(17.3)  | 2143(22.8) | 314(20.5)  |
| **Stage 2 Hypertension**      | 8594(7.5)   | 6351(7.3)   | 1147(7.5)   | 146(7.3)   | 809(8.6)  | 141(9.2)   |
| **Stage 3 Hypertension**      | 2945(2.6)   | 2016(2.3)   | 589(3.9)    | 39(2)     | 236(2.5)  | 65(4.2)    |

Abbreviations: SD, standard deviation, CVD, cardiovascular disease, BMI, body mass index, SBP, systolic blood pressure, DBP, diastolic blood pressure

1Classification of blood pressure defined as Chinese guideline for prevention and treatment of hypertension: Normal, SBP<140 and DBP<90; Optimal, SBP<120 and DBP<80; Prehypertension, 120≤SBP<140 or 80≤DBP<90; Hypertension, SBP≥140 or DBP≥90; Stage 1 Hypertension, 140≤SBP<160 or 90≤DBP<100; Stage 2 Hypertension, 160≤SBP<180 or 100≤DBP<110; Stage 3 Hypertension, SBP≥180 or DBP≥110.