# Study profile of The Xinjiang multi-ethnic cohort study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-048242 |
| Article Type: | Cohort profile |
| Date Submitted by the Author: | 23-Jan-2021 |
| Complete List of Authors: | Tian, Tian; Xinjiang Medical University, Xinjiang Medical University Tao, Luo Liu, Lirong; Xinjiang Medical University, Xinjiang Medical University Zhang, Zewen; Xinjiang Medical University, Xinjiang Medical University Sun, Qi Sun, Gaofeng Dai, Jianghong; Xinjiang Medical University, Xinjiang Medical University Yan, Hong; Xi An Jiao Tong Univ |
| Keywords: | EPIDEMIOLOGY, PUBLIC HEALTH, Diabetes & endocrinology < INTERNAL MEDICINE |
| | |

SCHOLARONE™
Manuscripts

**BMJ**

1 **Study profile of The Xinjiang multi-ethnic cohort study**

2 **Short Title:** Profile of XMC Study

3 Tian Tian[1#], Luo Tao[1#], Liu Lirong[1], Zhang Zewen[1], Sun Qi[2], Sun Gaofeng[3], Jianghong

4 Dai [1*], Yan Hong[4]

5 [1]Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang

6 Medical University, Urumqi, 830001, China;

7 [2]Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous Region,

8 Urumqi, 830099, China;

9 [3] Urumqi Municipal Center for Disease Control and Prevention, Urumqi, 830000,

10 China;

11 [4] Department of Epidemiology and Health Statistics, School of Public Health of Xi'an

12 Jiaotong University Health Science Center, Xi'an, 710061, China;

13 [#]These authors contributed equally

14 [*]Correspondence:

15 Jianghong Dai

16 Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang

17 Medical University, Urumqi, 830001, China

18 Email: epidjh@163.com

19 **Abstract**

20 **Purpose:** To investigate the causal link between diet and other lifestyle factors with

21 long-term health consequences, we established the Xinjiang multi-ethnic cohort study

22 (XMC), the first large-scale prospective cohort in Xinjiang, China.

23  **Participants**: XMC commenced in 2018 and enrolled participants from three study

24  sites (Urumqi, Hotan and Ili) in Xinjiang, China. Data collected include dietary data,

25  physical measurements, blood samples and urine samples. About one third of

26  participant were assessed habitual diet by interviewer-administered semi-quantitative

27  food frequency questionnaire (FFQ) which included 127 foods items at baseline

28  **Findings to date:** A total of 30,949 participants, with 32.37% from Urumqi, 41.75%

29  from Hotan and 25.88% from Ili were recruited in XMC. The average age of

30  participants were 56.21 years for men, and 54.75 years for women. More than 60% of

31  participants in all three survey sites reported an average consumption of fruit and

32  vegetable three or more times per week. In Hotan and Ili, the staple food was wheaten

33  food, whereas in Urumqi, rice and wheaten food were the food staples. Consumption

34  of white meat, such as fish and poultry, was lower in the three survey locations. The

35  five most common chronic diseases among participants across all three survey

36  locations were dyslipidemia, hypertension, cholecystitis, diabetes, ischaemic heart

37  disease.

38  **Future plans:** We will take advantage of the existing monitoring systems or database

39  of Xinjiang, China to obtain information on health outcomes (such as morbidity and

40  death events) for all participants of the cohort study. Repeated cross-sectional surveys

41  of study participants will be conducted on a bi-annual basis with surveys focusing on

42  research outcome being the primary concern.

43  **Key words:** Cohort Profile; Dietary habit; Chronic diseases; Longitudinal Cohort

44      **Strengths and limitations of this study**

45 • The Xinjiang multi-ethnic cohort is the first population cohort study of its kind
46   established in Xinjiang, China. The study includes a study population of more
47   than 30,000 people, of which about one third are urban residents, one third are
48   farmers, and the remaining third are animal herdsmen, which is broadly
49   representative of the demographics of residents in Xinjiang.

50 • This research commenced in 2018 and participants were surveyed once a year
51   until the end of the four-year research period (2018-2022). Data collect from
52   study participants (disease, death, migration, etc.) were verified by comparison
53   with the disease and death monitoring data network of the health and family
54   planning department, the medical record data of the hospital, the resident medical
55   insurance system and the medical institution, as well as annual household
56   registration and death registration data of the public security department. In the
57   final year of the cohort study, 10% of subjects were randomly selected for
58   telephone or face-to-face follow-up. After research period ended, we will keep
59   survey the subjects by a comprehensive physical examination project within
60   Xinjiang.

61 • This study collected information on the demographic characteristics of the study
62   participants by way of a survey questionnaire, as well as blood samples. Some
63   research subjects also retained genetic material such as RNA. About two-thirds of
64   the subjects of this study are ethnic minorities. Their ethnic background allows us
65   to compare health outcomes and health determinants of populations living in
66   distinct geographic locations. The results of this study found differences in the
67   dietary habits of people from different ethnic backgrounds which allows us to
68   hypothesize about associations between eating patterns and health outcomes.

69 • The study data are not freely available, but specific proposals for future
70   collaborations are welcome. Address to the research leader of Xinjiang multi-
71   ethnic cohort study (Dai Jianghong, Email: epdjh@163.com).

72 **Introduction**

73   Over the past several decades, China has established several large-scale

74    prospective population cohort studies. These have included the China Kadoorie

75    Biobank (CKB) study[1], a population study of 500,000 people among ten provinces in

76    China, and the Shandong Multi-center Longitudinal Cohort for Health Management[2],

77    which included a research cohort of 100 million people and covered a study period of

78    12 years. The Taizhou Longitudinal Study (TZL)[3] and Tianjin Chronic Low-Grade

79    Systemic Inflammation and Health Cohort Study (TCLSIH)[4] have also been among

80    the longitudinal cohort studies established during the same time period. These cohort

81    studies have focused on analyzing the interactions between health-related risk factors,

82    as well as environmental and genetic factors, with population health in China[5]. The

83    main research outputs of these cohort studies have had a substantial impact on public

84    health promotion in China, however, to date no prospective cohort studies have been

85    carried out in the Xinjiang Uygur Autonomous Region (Xinjiang) (Figure 1).

86         Xinjiang is the largest provincial administrative region in China, accounting for

87    one sixth of China's land mass (1.66 million square kilometers). The Altai Mountains

88    in the north of Xinjiang the Kunlun Mountains in the south and the Tianshan

89    Mountains in the middle divide the region into two distinct geographic areas; the

90    Tarim Basin in the South and the Junggar Basin in the north. The topography of

91    Xinjiang is complex with geographical features varying from mountains and basins to

92    grasslands, deserts and oases. Xinjiang has a multi-ethnic population with

93    approximately 20 million people (about 60% of the total population of Xinjiang)

94    belonging to Uyghur, Kazakh, Hui and other non-Han ethnic groups[4, 6] and a unique

95    diet which may be correlated with health outcomes in the region.

96    The main risk factors associated with non-communicable chronic diseases are

97    modifiable risk factors such as poor diet, lack of physical activity, and alcohol and

98    tobacco use. High-quality epidemiological data on diet, environmental and genetic

99    determinants of non-communicable chronic diseases and long-term outcomes are

100   essential for developing public health strategies to reduce the burden of non-

101   communicable diseases. Results of statistical analyses carried out on surveillance data

102   found that heart disease, malignant tumors, and cardiovascular disease were the main

103   contributors to the overall burden of chronic health conditions in 2015, and accounted

104   for 77.67% of all deaths in Xinjiang.[7]

105   While extensive research the etiology of chronic diseases has been carried out on

106   across much of mainland China, the risk factors underpinning chronic disease may

107   vary between populations, and substantial uncertainty exists as to how important these

108   risk factors are in different settings. Because of the unique geographic environment,

109   genetic background and population make-up of Xinjiang, the etiologies of chronic

110   disease deduced from other cohort studies carried out in China may therefore not be

111   fully applicable to the population of this region. The unique geographic environment,

112   genetic background and population make-up also leads to the special eating habits of

113   the residents in Xinjiang. The main objectives of this study were to identify diet risk

114   factors and other determinants of non-communicable diseases among the population

115   of Xinjiang.

116   **Material and Methods**

117   *Study design and Organization*

118     The study was conducted in collaboration with the Xinjiang Medical University,

119     the Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous

120     Region and the Urumqi Municipal Center for Disease Control and Prevention, and

121     was completed under the guidance of Xi'an Jiaotong University. Ethical approval for

122     this study was granted by the Ethics Committee of Academy of Traditional Chinese

123     Medicine in Xinjiang Uygur Autonomous Region (2018XE0108).

124     The locations selected for inclusion in the cohort study, Urumqi, Hotan and Ili,

125     were chosen in accordance with the geographical distribution of minority populations

126     in the Xinjiang Uygur Autonomous Region (Figure 2). Urumqi city is the capital of

127     the Xinjiang and is the cultural and political center of the region. Surveys on factors

128     such as health conditions, health-related behaviors, and diet were carried out in 20

129     community health service centers in Urumqi in order to be representative of the urban

130     population of Xinjiang. In Hotan, surveys were carried out in local villages where

131     more than 95% of residents are of the Uyghur ethnicity. The highest level of

132     educational attainment among this population is generally low and farming is the

133     predominant occupation in this area. In addition to this, the population living in Hotan

134     have distinctive dietary habits, the health outcomes of which may be possible to

135     establish by comparing diet and health outcomes between this population and

136     populations outside of this survey location. In Ili, surveys were conducted in

137     townships where the distribution of Han, Uyghur, Kazak and Hui is relatively

138     balanced. Most study participants recruited in this area were animal herders, the

139     survey results therefore being reflective of health status and health-related risk factors

140  of herdsmen in rural areas of Xinjiang. These diverse survey sites were selected in

141  order to be reflective of the health status of both urban and rural residents, workers or

142  farmers and herders, and different ethnicities in Xinjiang.

143  *Study participants*

144  Eligibility criteria for inclusion in the cohort study were adults aged 35 to 74

145  (born between 1943 and 1982), without any physical or communication disabilities,

146  with the ability to formally consent to participation in the study cohort, residing

147  permanently (at least 1 year of residency) in study sites, and whose disease incidence

148  data belonged to the local department of health. We excluded migratory populations

149  and temporary residents, as these participants are more likely to be lost to follow-up.

150  We also excluded armed forces and staff members stationed in the investigation site,

151  as long-term follow up is also challenging among this population. A total of 30949

152  individuals were included in the cohort study with 10017 participants from Urumqi,

153  12921 participants from Hetian land and 8011 participants enrolled in Yili. All

154  participants gave written informed consent before the study began.

155  *Patient and public involvement*

156  No patient involved.

157  *Follow-up survey*

158  Beginning in September 2016, Xinjiang, the largest autonomous region in China,

159  launched a comprehensive physical examination project within the region. All

160  residents in Xinjiang can participate in this free annual physical examination, with the

161  cost covered by the local government. The residents were recruited by government

162 announcement. The project will continue for decades. During the past four years,

163 Xinjiang has set up more than 2,800 physical examination centers, and invested a total

164 of 3.90 billion yuan in national physical examination. By the end of September 2018,

165 a total of 51.03 million subjects participated in this free annual physical examination.

166 Results of physical examination have formed a complete longitudinal database, after

167 our cohort fund project ends, we can link with the database through ID and continue

168 to follow up the research subjects. We have matched the baseline survey data of XMC

169 subjects with the personnel database of the national physical examination within

170 Xinjiang in 2019, more than 70% of subjects in XMC participated in the national

171 physical examination within Xinjiang in 2019.

172     This free annual physical examination for all residents in Xinjiang, which

173 included but were not limited to, physical examination, health-risk survey,

174 biochemical blood examination, ultrasound and X-ray were carried out by the

175 Xinjiang Government. These health checks are routinely attended by more than 80%

176 of adults residing in the region, allowing health data for all study participants to be

177 collected annually. Participant follow-up was facilitated by recording of a unique

178 individual identification (ID) number associated with everyone across several

179 common monitoring or data systems. These data systems include medical record

180 information systems obtained from hospitals, medical insurance management systems,

181 maternal and child health information systems, public security household registration

182 management systems, civil affairs (burial) management, disease prevention and

183 control systems, disease incidence monitoring and systems which monitor causes of

184    death in the region. Data from these different sources can be used concordantly for

185    mutual review and for supplementing survey data.

186        In the final year of the cohort study, 10% of subjects were randomly selected for

187    telephone or face-to-face follow-up. Study participants given priority for this follow

188    up included those that had not had any measurable outcome during the study period.

189    *Data and blood samples collection*

190        The data collection methods applied for the Xinjiang multi-ethnic cohort study

191    included a questionnaire survey, physical examination, and collection of biological

192    samples; measures which are essential for any longitudinal study.

193        The questionnaire collected information on sociodemographics, tea and coffee

194    consumption, alcohol intake, tobacco use, dietary habits, passive smoking and indoor

195    air pollution, personal and family medical history, physical activity, mental health and

196    reproductive history (females study participants only; Table 1). The survey

197    questionnaire design included different levels of detail for collecting data on dietary

198    habits in the three different survey locations. In Ili, the data collected on dietary habits

199    included specific information as to the types of food consumed, as well as the

200    frequency and intake of specific food types. For example, in Ili, information on the

201    frequency and intake amount of fruit consumed, as well as specifics pertaining to fruit

202    type, i.e., number of apples, bananas and oranges consumed, was collected. In the

203    other study settings however, only information on frequency and total fruit intake was

204    collected.

| Demographic data | Name, Gender, ID card, Medical insurance, Date of birth, Education level, Marital status, Occupation, Financial income |
|---|---|

| | |
|---|---|
| Tea and coffee consumption | Frequency of tea drinking in the past year, Previous tea drinking habits, Types of tea drinking, Frequency of coffee consumption |
| Alcohol intake | Frequency and amount of drinking in the past year, previous drinking habits (drinking refers to drinking in the past year under normal circumstances, not holidays or special periods, such as during marriage) |
| Tobacco use | Frequency and amount of smoking in the past year, previous smoking habits (current smoker defined as daily or almost daily smoking; former smoker defined as stopping smoking for at least 6 months; non-smoker defined as that the cumulative amount of smoking in lifetime does not exceed 100) |
| Dietary habits | Intake of 30 common foods, including staple foods, animal and plant foods, soy products, dairy products, vegetables, and fruits; Types of daily consumption oil; Use of nutrient supplements; Spicy food and vinegar consumption; Household refrigerator use time |
| Passive smoking and indoor air pollution | Whether had history of live with current smoker; Frequency of weekly exposure to passive smoking and cumulative exposure time; Cooking, heating, household fuel use and house decoration situation; Occupational exposure history of air pollutants |
| Personal and family medical history | Self-health evaluation; chest and respiratory symptoms; history of 18 common diseases including diabetes, acute myocardial infarction, hypertension, asthma, and malignant tumors; history of blood transfusion and constipation; history of 5 common chronic diseases among family members |
| Physical activity | Work-related physical activity in the past year; way of commuting and time spent on commuting; amateur physical exercise; housework activities; frequency and length of physical activity (i.e., sweating, heartbeat, etc.); whether had slimed in the past year; weight change |
| Mental health | Satisfaction with current living conditions; 10 major events that may have a serious impact on life in the past two years; sleep conditions (including insomnia, daily sleep time, whether to snoring while sleeping, etc.); depression, anxiety disorders and unexplained phobias; quality of life |
| Reproductive history (only for female) | Menstrual history of female (age at menopause and menarche), pregnancy, fertility, breastfeeding, contraceptive use, and surgery |

205

206     Data collected during the physical examination included height (using a medical

207     height gauge with an accuracy of 0.1cm), weight (using a medical electronic scale

208     with an accuracy of 0.1kg), waist circumference (using a soft measuring tape with an

209     accuracy of 0.1 cm; the lower edge of the measuring tape was placed at the highest

210     points on both sides of the hip bone, and horizontal measurement of the waist

211     circumference taken), heart rate and blood pressure (measured using a medical arm

212     electronic blood pressure monitor with an accuracy of 2 mmHg; two measurements

213     per study participant were taken, following a rest period of five to ten minutes), body

214     fat composition (measured using a body composition analyzer, TANITA DC-

215     430MA ). Participants were asked to remove jacket, shoes and hat while all physical

216     measurements were being taken.

217     A vacuum blood collection device with intravenous anticoagulant was used to

218     collect a 20 ml blood sample for each participant. A 4 ml blood sample was used for

219     biochemical examination and routine blood examination. These examinations were

220     completed at the nearest township health service center to the survey location. Whole

221     blood samples (3 ml) were transferred to three cryopreservation tubes immediately

222     after blood samples were collected. Blood samples used to separate plasma and white

223     blood cells were centrifuged within two hours of blood sample collection (4℃ at

224     3000 rpm for 10 minutes). Samples were stored at -196℃ in liquid nitrogen

225     containers before and after bi-monthly transportation to Urumqi. RNA protection

226     solution was also added to some samples for subsequent RNA detection. All the blood

227     and urine samples are placed in a special biological sample bank, which is managed

228    by a specially-assigned person using an electronic management system.

229        For a large study such as this to be practicable and economically feasible in a

230    resource-poor setting, it was necessary for study procedures to be simple and

231    streamlined. In this study, the full assessment carried out at recruitment, including

232    obtaining informed consent, administering the questionnaire survey, carrying out the

233    physical examination and obtaining blood samples, took an average of 60 minutes to

234    complete. Investigators from all three survey locations were trained by the same

235    junior investigator and postgraduate of the Xinjiang Medical University. The survey

236    equipment was also consistent across all three survey locations and equipment

237    calibration was carried out every day.

238    *Statistical analyses*

239        For the baseline profile, descriptive statistics were calculated for baseline data

240    regarding demographic data, tea and coffee consumption, alcohol intake, tobacco use,

241    dietary habits, passive smoking and indoor air pollution, personal and family medical

242    history, physical activity, mental health, reproductive history (only for female). All

243    analyses were conducted using SAS version 9.4.

244    **Results**

245        A total of 30,949 participants were recruited as part of the Xinjiang multi-ethnic

246    cohort study form Urumqi, Hotan and Ili in Xinjiang, China. The number of

247    participants recruited from Urumqi, Hotan and Ili was 10017, 12921 and 8011,

248    respectively (Table 2). The average age of study participants was 56.21 years for men,

249    and 54.75 years for women. In Urumqi, more than half of the subjects were aged over

250    65 years old, significantly higher than the other two regions. The proportion of Han

251    Chinese in the Urumqi study population was close to 90%, while the proportion of

252    Uyghur in the Hotan study population was over 99%. The proportion of ethnic groups

253    represented in the study population recruited in the Ili region, conversely, was

254    relatively balanced, indicating that ethnic composition varies significantly from region

255    to region. The average level of highest educational attainment, and average income

256    among the Urumqi population was significantly higher than that of the Hotan and Ili

257    populations. The proportion of people who had been educated to primary or pre-

258    primary level in Hotan and Yili were 87.65% and 71.14%, respectively.

259 Table 2. Demographic characteristics of the study participants by site

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Men | Women | Total | Men | Women | Total | Men | Women | Total |
| Age(years) | | | | | | | | | |
| <45 | 231(5.87) | 330(5.43) | 561(5.60) | 1040(21.17) | 2337(29.18) | 3377(26.14) | 1120(29.60) | 1250(29.57) | 2370(29.58) |
| 45-54 | 570(14.47) | 1021(16.80) | 1591(15.88) | 1408(28.66) | 2871(35.85) | 4279(33.12) | 1248(32.98) | 1537(36.36) | 2785(34.76) |
| 55-64 | 824(20.92) | 1358(22.34) | 2182(21.78) | 1364(27.76) | 1848(23.08) | 3212(24.86) | 870(22.99) | 994(23.52) | 1864(23.27) |
| 65- | 2313(58.74) | 3370(55.44) | 5683(56.73) | 1101(22.41) | 952(11.89) | 2053(15.89) | 446(14.43) | 446(10.55) | 992(12.38) |
| Race | | | | | | | | | |
| Han | 3457(87.79) | 5432(89.36) | 8889(88.74) | 6(0.12) | 6(0.07) | 12(0.09) | 572(15.12) | 522(12.35) | 1094(13.66) |
| Hui | 310(7.87) | 392(6.45) | 702(7.01) | 4(0.08) | 15(0.19) | 19(0.15) | 1336(35.31) | 1333(31.54) | 2669(33.32) |
| Uyghur | 105(2.67) | 168(2.76) | 273(2.73) | 4897(99.67) | 7969(99.51) | 12866(99.57) | 908(24.00) | 1361(32.20) | 2269(28.32) |
| Kazakh | 41(1.04) | 53(0.87) | 94(0.94) | 6(0.12) | 13(0.16) | 19(0.15) | 883(23.34) | 951(22.50) | 1834(22.89) |
| Other | 25(0.63) | 34(0.56) | 59(0.59) | NA | 5(0.06) | 5(0.04) | 85(2.25) | 60(1.42) | 145(1.81) |
| Education | | | | | | | | | |
| Primary/less | 1064(27.02) | 2534(41.68) | 3598(35.92) | 4093(83.31) | 7232(90.31) | 11325(87.65) | 2582(68.23) | 3117(73.74) | 5699(71.14) |
| Secondary | 2320(58.91) | 3020(49.68) | 5340(53.31) | 774(15.75) | 734(9.17) | 1508(11.67) | 1166(30.81) | 1071(25.34) | 2237(27.92) |
| Degree or above | 553(14.04) | 517(8.50) | 1070(10.68) | 32(0.65) | 18(0.22) | 50(0.39) | 28(0.74) | 26(0.62) | 54(0.67) |
| Refused/missing | 1(0.03) | 8(0.13) | 9(0.09) | 14(0.28) | 24(0.30) | 38(0.29) | 8(0.21) | 13(0.31) | 21(0.26) |

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Men | Women | Total | Men | Women | Total | Men | Women | Total |
| **Marriage** | | | | | | | | | |
| Married | 3696(93.85) | 5045(82.99) | 8741(87.26) | 4319(87.91) | 6165(76.99) | 10484(81.14) | 3500(92.49) | 3586(84.84) | 7086(88.45) |
| Widowed/separated | 197(5.00) | 1006(16.55) | 1203(12.01) | 538(10.95) | 1773(22.14) | 2311(17.89) | 227(6.00) | 614(14.53) | 841(10.50) |
| Never married | 42(1.07) | 19(0.31) | 61(0.61) | 19(0.39) | 6(0.07) | 25(0.19) | 43(1.14) | 6(0.14) | 49(0.61) |
| Refused/missing | 3(0.08) | 9(0.15) | 12(0.12) | 37(0.75) | 64(0.80) | 101(0.78) | 14(0.37) | 21(0.50) | 35(0.44) |
| **Income (Yuan)** | | | | | | | | | |
| under 10000 | 632(16.05) | 912(15.00) | 1544(15.41) | 2689(54.73) | 5139(64.17) | 7828(60.58) | 414(10.94) | 616(14.57) | 1030(12.86) |
| 10000-19999 | 412(10.46) | 585(9.62) | 997(9.95) | 1196(24.34) | 1649(20.59) | 2845(22.02) | 1311(34.65) | 1712(40.50) | 3023(37.74) |
| 20000-34999 | 411(10.44) | 737(12.12) | 1148(11.46) | 453(9.22) | 501(6.26) | 954(7.38) | 1157(30.58) | 1093(25.86) | 2250(28.09) |
| 35000-49999 | 504(12.80) | 885(14.56) | 1389(13.87) | 148(3.01) | 198(2.47) | 346(2.68) | 393(10.39) | 339(8.02) | 732(9.14) |
| 50000 or above | 1978(50.23) | 2952(48.56) | 4930(49.22) | 416(8.47) | 500(6.24) | 916(7.09) | 503(13.29) | 459(10.86) | 962(12.01) |
| Refused/missing | 1(0.03) | 8(0.13) | 9(0.09) | 11(0.22) | 21(0.26) | 32(0.25) | 6(0.16) | 8(0.19) | 14(0.17) |
| **Anthropometric index** | | | | | | | | | |
| Height | 169.03±6.26 | 157.98±19.33 | 162.33±16.47 | 163.32±7.29 | 153.60±6.81 | 157.30±8.44 | 167.13±6.98 | 155.98±7.05 | 161.24±8.95 |
| Weight | 73.82±10.69 | 63.10±11.46 | 67.31±12.33 | 66.71±12.38 | 59.63±12.04 | 62.32±12.64 | 72.24±12.06 | 65.20±12.33 | 68.52±12.69 |
| Waist circumference | 92.40±17.00 | 87.09±15.47 | 89.18±16.30 | 91.88±11.50 | 91.44±12.29 | 91.61±12.00 | 90.56±11.02 | 88.49±11.80 | 89.46±11.48 |

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Men | Women | Total | Men | Women | Total | Men | Women | Total |
| Body mass index | 25.81±3.31 | 25.35±4.42 | 25.53±4.02 | 24.96±4.04 | 25.24±4.67 | 25.14±4.44 | 25.82±3.75 | 26.76±4.51 | 26.31±4.20 |

260

261     Approximately 50% of participants surveyed in Urumqi had an average annual

262     income of more than 50,000, whereas about 60% of subjects in Hotan had an average

263     annual income of less than 10,000. Study participants in Urumqi were found to have

264     the highest average height, while study participants in Ili were found to have the

265     highest average weight, with study participants in Hotan having the highest average

266     waist circumference. The average body mass index of the subjects in the three regions

267     exceeded the diagnostic threshold for obesity in China, with study participants in Ili

268     region having the highest average body mass index.

269     The five most common chronic diseases among participants across all three

270     survey sites were dyslipidemia, hypertension, cholecystitis, diabetes, ischaemic heart

271     disease with prevalence of 34.55%, 32.66%, 14.34%, 10.07%, and 8.74%,

272     respectively (Table 3).

273     Table 3. Self-reported disease prevalence among subjects in different sites

| Diseases | Urumqi | | Hotan | | Ili | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Crude | Adjusted | Crude | Adjusted | Crude | Adjusted | Crude | Adjusted |
| Cardiovascular diseases | | | | | | | | |
| Hypertension* | 47.19 | 29.29 | 37.00 | 32.64 | 38.61 | 36.09 | 40.74 | 32.66 |
| Hypertension | 32.64 | 17.87 | 27.35 | 23.76 | 16.71 | 15.23 | 26.31 | 19.80 |
| IHD | 8.18 | 3.28 | 13.54 | 12.15 | 8.44 | 7.72 | 10.48 | 8.74 |
| Other heart disease | 0.57 | 0.32 | 2.63 | 2.34 | 0.69 | 0.62 | 1.46 | 1.27 |
| Stroke | 3.43 | 1.62 | 4.91 | 4.31 | 1.61 | 1.47 | 3.58 | 2.82 |
| Metabolic disease and kidney diseases | | | | | | | | |
| Diabetes* | 23.93 | 19.14 | 7.26 | 6.74 | 9.80 | 9.19 | 13.11 | 10.07 |
| Diabetes | 14.73 | 8.81 | 4.25 | 3.87 | 4.69 | 4.36 | 7.76 | 5.54 |
| Dyslipidemia* | 35.00 | 35.08 | 33.53 | 34.05 | 34.83 | 35.08 | 34.32 | 34.55 |
| CKD | 0.68 | 0.56 | 6.20 | 6.28 | 0.99 | 0.93 | 3.06 | 3.17 |
| Anemia* | 3.31 | 5.30 | 5.89 | 5.55 | 5.30 | 5.26 | 4.98 | 5.44 |
| Chronic respiratory diseases | | | | | | | | |
| Chronic bronchitis | 4.48 | 2.16 | 15.89 | 14.17 | 3.31 | 3.12 | 8.94 | 7.87 |
| Emphysema | 0.29 | 0.15 | 3.00 | 2.57 | 0.51 | 0.49 | 1.48 | 1.25 |
| COPD | 0.40 | 0.17 | 1.22 | 1.14 | 0.20 | 0.18 | 0.69 | 0.62 |
| Asthma | 1.11 | 0.65 | 2.86 | 2.45 | 0.22 | 0.20 | 1.61 | 1.36 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tuberculosis | 0.69 | 0.57 | 2.84 | 2.55 | 0.54 | 0.50 | 1.55 | 1.37 |
| Digestive diseases | | | | | | | | |
| Chronic hepatitis | 0.49 | 0.53 | 4.85 | 4.91 | 1.34 | 1.33 | 2.53 | 2.82 |
| Cholecystitis | 6.50 | 3.12 | 23.28 | 21.28 | 13.66 | 12.49 | 15.36 | 14.34 |
| Peptic ulcer | 0.91 | 0.74 | 6.61 | 6.33 | 2.11 | 2.01 | 3.60 | 3.61 |
| Skeletal disorders | | | | | | | | |
| Osteoporosis | 2.15 | 0.83 | 7.48 | 6.50 | 3.74 | 3.52 | 4.79 | 4.21 |
| Fracture | 1.93 | 1.24 | 5.82 | 6.10 | 5.75 | 5.87 | 4.54 | 4.96 |
| Cancer | 0.63 | 0.32 | 0.67 | 0.54 | 0.46 | 0.43 | 0.60 | 0.48 |

274    *Self-reported combined with biological detection

275    Adjusted: Age and gender adjusted based on 2010 China census data

276    IHD: Ischaemic heart disease, Other heart disease: included Rheumatic heart disease and Cor pulmonale,

277    CKD: Chronic kidney disease

278        Disease prevalence varies substantially across the three survey locations, which

279    may be attributed to differences in nationality, diet, education, socioeconomics and

280    living habits between the three areas. After standardizing for age and sex based on

281    2010 China census data, the prevalence of hypertension in Ili was 36.09%, higher than

282    that in Urumqi (29.29%) and Hotan (32.64%). According to a survey carried out

283    between 2012 to 2015 the average prevalence of hypertension in China was 23%[8].

284    The prevalence of hypertension in the XMC study sites, as determined by this study,

285    was much higher than the national average. Study participants in Urumqi had a higher

286    prevalence of diabetes (19.14%) compared with those in Ili and Hotan. The results of

287    this study also found that the prevalence of diabetes in Ili and Hotan was slightly

288    lower than the national average (10.9%).[9] Urumqi has a high prevalence of diabetes

289    and hypertension, which may be associated with participants residing in urban

290    communities, compared with rural dwelling populations in Hotan and Ili. In addition

291    to this, populations living in Urumqi have a higher average income and a lower

292    proportion of people belonging to the physical labor workforce than other two

293    regions. The prevalence of chronic kidney disease among participants in Hotan

294    (6.28%) was approximately ten times higher than that of Urumqi (0.56%) and six

295    times that of Ili (0.93%). Prevalence of cholecystitis in Hotan (21.28%) was

296    approximately seven times that of the prevalence in Urumqi (3.12%). Results of this

297    study also found that the prevalence of COPD in Xinjiang (0.62%) was much lower

298    than the national level (3.84%)[10]. Regarding Cholecystitis, the data on national

299    average were limited. A previous epidemiological survey showed that the incidence of

300    gallstones in Xinjiang was 11.83%.[11] The prevalence of Cholecystitis in Ili (12.49%)

301    and Hotan (21.28%) was much higher than previous recorded levels. This may be

302    associated with Hotan's local dietary habits. The prevalence of ischemic heart disease,

303    stroke, chronic hepatitis and CKD in Hotan was also found to be higher than that of

304    the other two sites.

305        At baseline, more than 60% of participants in all three survey locations reported

306    an average consumption of fruit and vegetable three or more times per week. In Hotan

307    and Ili, the staple food was wheaten food, whereas in Urumqi, rice and wheaten food

308    were the food staples. Dietary habits in Hotan, which has a predominantly Uyghur

309    population, consisted predominantly of mutton, with more than 70% consuming

310    mutton three or more times per week. The proportion of the population who

311    consumed pork, mutton or beef three or more times per week in Urumqi by contrast,

312    where the population is predominantly Han, was 28.26%, 20.24 and 24.28%,

313    respectively. Consumption of white meat, such as fish and poultry, was lower in the

314    three survey locations. Fresh meat consumed by Xinjiang residents consisted mainly

315    of red meat such as beef and mutton. Some studies have suggested that too high an

316  intake of red meat might increase the risk of metabolic syndrome [12,13]. The study also

317  showed that consumption of fresh milk was associated with a lower risk of Metabolic

318  Syndrome[13]. In this study, the proportion of people drinking milk three or more times

319  per week was lower in Ili and particularly in Hotan (<10% participants surveyed).

320  More than 60% of people in Ili reported drinking milk tea ≥3 times a week, however,

321  with few of them drinking fresh milk directly.

322      A diet high in salt and fat are associated with increased risk of high blood

323  pressure and ischemic heart disease[14]. The consumption of coarse grains, eggs, milk

324  and yogurt three or more times per week among people surveyed in Urumqi was

325  higher than among people surveyed in Ili and Hotan. In addition to this, consumption

326  of fowl, fish or seafood, and beans three or more times per week was low in all three

327  survey locations (<10% of participants surveyed). The results of the survey on dietary

328  habits reported here are generally consistent with a recent survey[15], which found that

329  consumption of cereals, meat and oil in Xinjiang exceeded maximum recommended

330  allowances, while consumption of eggs, seafood products, dairy products, nuts and

331  legumes were below the minimum recommended allowances. Notably, this study

332  reported a higher consumption of fruit and vegetable in all three regions than previous

333  studies on fruit and vegetable consumption in Xinjiang[13,16].

334  **Discussion**

335      This is the largest prospective study that has been undertaken to date in Xinjiang,

336  China. The main strength of the Xinjiang multi-ethnic cohort study is the inclusion of

337  over 20,000 survey participants from diverse ethnic backgrounds such as Uyghur,

338    Kazakh, and Hui, which allows prevalence of chronic conditions, as well as dietary

339    habits, various health determinants and socio-demographics between distinct

340    populations to be compared. It is already well established that the dietary habits of

341    Uyghur, Kazakh, and Hui differ from those of Han Chinese. We have investigated the

342    dietary habits of all the study participants as part of this cohort study, in conjunction

343    with prevalence of non-communicable diseases to compare health determinants and

344    prevalence of chronic conditions between these three survey locations.

345         Another important strength of this study was the survey questionnaire, which

346    collected details on major risk factors such as alcohol intake, tobacco use (we

347    investigated current as well as past usage quantity, frequency, type), mental health and

348    quality of life. The three survey locations selected for this study were broadly

349    representative of the different populations residing in Xinjiang, including animal

350    herders, and urban and rural residents. The results of survey conducted here can be

351    used to compare differences in prevalence and type of non-communicable chronic

352    diseases among these three populations.

353         There were also several limitations with the Xinjiang multi-ethnic cohort study.

354    Firstly, we did not use random sampling to select study participants, but rather a

355    convenience sampling strategy whereby eligible study participants from

356    predetermined survey locations were recruited until an adequate sample size had been

357    met. As this study is the first large-scale cohort study to have been carried out in

358    Xinjiang, and data collected as part of previous cross-sectional health surveys of

359    Xinjiang residents have not yet been published, we cannot compare differences in the

360  distribution of diseases and related factors between our cohort study population and

361  the wider Xinjiang population. Another limitation is that diseases were self-reported,

362  although a detailed and comprehensive questionnaire included a screening for self-

363  assessment of health and quality of life was carried out. As a result, some results are

364  subject to reporting and recall bias.

365      We will take advantage of the existing monitoring systems or database of

366  Xinjiang, China to obtain information on health outcomes (such as morbidity and

367  death events) for all participants of the cohort study. Repeated cross-sectional surveys

368  of study participants will be conducted on a bi-annual basis with surveys focusing on

369  research outcome being the primary concern. Data will be obtained by administering

370  questionnaires used for baseline research, as well as additional information on health

371  determinants. Although the study population is relatively large, the availability of

372  repeated measures, extensive biobank blood and urine samples, and health

373  information systems linked by an ID number will allow this study to assess within-

374  and between-person variability over time in major risk factors among the population

375  of people living in Xinjiang, China.

376      Baseline data from the Xinjiang multi-ethnic cohort study are put through a

377  process of editing. The study data are not freely available, but specific proposals for

378  future collaborations are welcome. Address to the research leader of Xinjiang multi-

379  ethnic cohort study.

380  **Acknowledgment:** The authors are grateful to all the participants of the baseline

381  survey of the XMC Study and the staff at each site for their cooperation. The Xinjiang

382    multi-ethnic cohort study is funded by a project of the National Key Research and

383    Development Project of China (Grant number: SQ2017YFSF090013) and a sub-

384    project of the National Key Research and Development Project of China (Grant

385    number: 2017YFC0907203).

386    **Conflict of interest:** The authors declare no competing interests.

387    **References**

388    1.   Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey
389         methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;**40**(6):1652–
390         1666.

391    2.   Zhang T, Zhang C, Zhang Y, et al. Metabolic syndrome and its components as predictors of
392         nonalcoholic fatty liver disease in a northern urban Han Chinese population: a prospective
393         cohort study. *Atherosclerosis*. 2015;**240**(1):144–148.

394    3.   Wang X, Lu M, Qian J, et al. Rationales, design and recruitment of the Taizhou Longitudinal
395         Study. *BMC Public Health*. 2009;**9**:223.

396    4.   Song K, Du H, Zhang Q, et al. Serum immunoglobulin M concentration is positively related to
397         metabolic syndrome in an adult population: Tianjin Chronic Low-Grade Systemic
398         Inflammation and Health (TCLSIH) Cohort Study. *PLoS ONE*. 2014;**9**(2):e88701.

399    5.   Li L, Lyu J. Large prospective cohort studies: a review and update. *Zhonghua Liu Xing Bing*
400         *Xue Za Zhi*. 2015;**36**(11):1187–1189.

401    6.   He J, Guo S, Liu J, et al. Ethnic differences in prevalence of general obesity and abdominal
402         obesity among low-income rural Kazakh and Uyghur adults in far western China and
403         implications in preventive public health. *PLoS ONE*. 2014;**9**(9):e106723.

404    7.   ZHE W, PUERHATI W, LIU L, LIAO P, ZHANG Y, ZHANG R. Analysis of the death causes
405         of residents in 15 surveillance sites in Xinjiang in 2015. *Modern Preventive Medicine*.
406         2017;**44**(22):4186–4190.

407    8.   Wang Z, Chen Z, Zhang L, et al. Status of Hypertension in China: Results From the China
408         Hypertension Survey, 2012-2015. *Circulation*. 2018;**137**(22):2344–2356.

409    9.   Wang L, Gao P, Zhang M, et al. Prevalence and Ethnic Pattern of Diabetes and Prediabetes in
410         China in 2013. *JAMA*. 2017;**317**(24):2515–2523.

411    10.  Chan KY, Li X, Chen W, et al. Prevalence of chronic obstructive pulmonary disease (COPD)
412         in China in 1990 and 2010. *J Glob Health*. 2017;**7**(2):020704.

413    11.  Zhu L, Aili A, Zhang C, Saiding A, Abudureyimu K. Prevalence of and risk factors for
414         gallstones in Uighur and Han Chinese. *World J Gastroenterol*. 2014;**20**(40):14942–14949.

415    12.  Oliveira Otto MC de, Alonso A, Lee D-H, et al. Dietary intakes of zinc and heme iron from
416         red meat, but not from other sources, are associated with greater risk of metabolic syndrome
417         and cardiovascular disease. *J Nutr*. 2012;**142**(3):526–533.

418    13.  Guo H, Gao X, Ma R, et al. Prevalence of Metabolic Syndrome and its Associated Factors
419         among Multi-ethnic Adults in Rural Areas in Xinjiang, China. *Sci Rep*. 2017;**7**(1):17643.

420    14.  Zhai FY, Du SF, Wang ZH, Zhang JG, Du WW, Popkin BM. Dynamics of the Chinese diet

421    and the role of urbanicity, 1991-2011. *Obes Rev*. 2014;**15 Suppl 1**:16–26.

422    15.  Yin J, Yang D, Zhang X, et al. Diet shift: Considering environment, health and food culture.

423         *Sci Total Environ*. 2020;**719**:137484.

424    16.  Zhai F, He Y, Wang Z, Hu Y. Status and characteristic of dietary intake of 12 minority

425         nationalities in China. *Wei Sheng Yan Jiu*. 2007;**36**(5):539–541.

426

427    Figure legends

428    Table 1. Questionnaire data collected in the Xinjiang multi-ethnic cohort study

429    Table 2. Demographic characteristics of the study participants by site

430    Table 3. Self-reported disease prevalence among subjects in different sites

431    Figure 1. Locations of on-going prospective cohorts in China

432    Figure 2. Map of Xinjiang, China, showing the locations of three surveyed sites (Red five-pointed star)

433    in the XMC

434    Figure 3. Percentage of common food intake in different sites in the XMC

435

Figure 1. Locations of on-going prospective cohorts in China

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Figure 2. Map of Xinjiang, China, showing the locations of three surveyed sites (Red five-pointed star) in the XMC

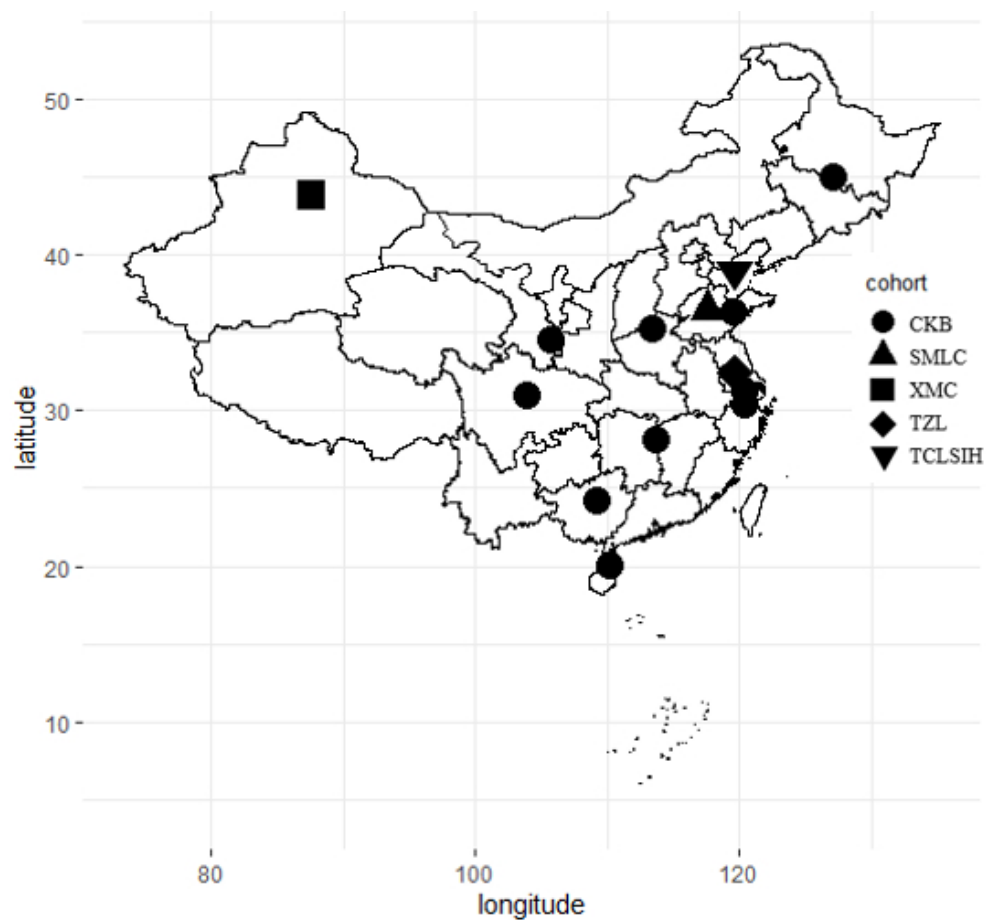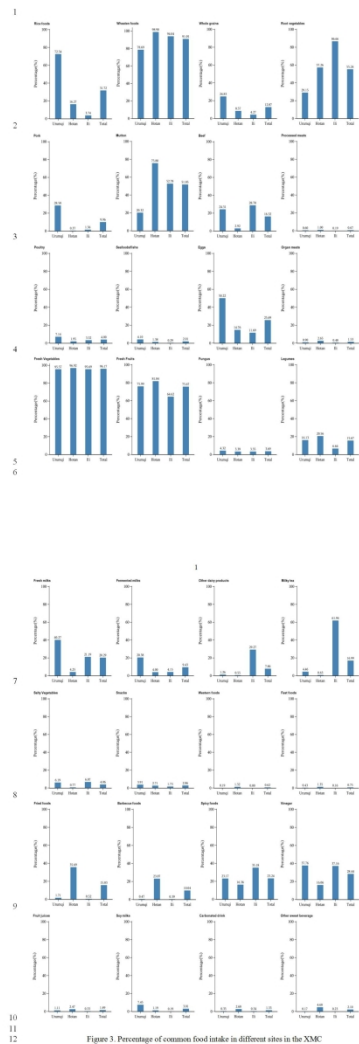Figure 3. Percentage of common food intake in different sites in the XMC

419x1063mm (96 x 96 DPI)

# BMJ Open

## Study profile of The Xinjiang multi-ethnic cohort study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-048242.R1 |
| Article Type: | Cohort profile |
| Date Submitted by the Author: | 01-Nov-2021 |
| Complete List of Authors: | Tao, Luo; Xinjiang Medical University, Department of Epidemiology and Biostatistics, School of Public Health<br>Tian, Tian; Sun Yat-Sen University, Shenzhen Campus of Sun Yat-sen University, School of Public Health (Shenzhen),<br>Liu, Lirong; Xinjiang Medical University, Department of Epidemiology and Biostatistics, School of Public Health<br>Zhang, Zewen; Xinjiang Medical University, Department of Epidemiology and Biostatistics, School of Public Health<br>Sun, Qi; Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous Region<br>Sun, Gaofeng; Urumqi Municipal Center for Disease Control and Prevention<br>Dai, Jianghong; Xinjiang Medical University, Xinjiang Medical University<br>Yan, Hong; Xi'an Jiaotong University |
| <b>Primary Subject Heading</b>: | Epidemiology |
| Secondary Subject Heading: | Public health |
| Keywords: | EPIDEMIOLOGY, PUBLIC HEALTH, Diabetes & endocrinology < INTERNAL MEDICINE |

SCHOLARONE™
Manuscripts

**BMJ**

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

1  **Study profile of The Xinjiang multi-ethnic cohort study**

2  **Short Title:** Profile of XMC Study

3  Luo Tao[1#], Tian Tian[2#], Liu Lirong[1], Zhang Zewen[1], Sun Qi[3], Sun Gaofeng[4], Jianghong

4  Dai [1*], Yan Hong[5]

5  [1]Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang

6  Medical University, Urumqi, 830001, China;

7  [2]Shenzhen Campus of Sun Yat-sen University, School of Public Health (Shenzhen),

8  Sun Yat-sen University, Shenzhen, Guangdong, 518107, China;

9  [3]Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous Region,

10  Urumqi, 830099, China;

11  [4]Urumqi Municipal Center for Disease Control and Prevention, Urumqi, 830000, China;

12  [5]Department of Epidemiology and Health Statistics, School of Public Health of Xi'an

13  Jiaotong University Health Science Center, Xi'an, 710061, China;

14  [#]These authors contributed equally

15  [*]Correspondence:

16  Jianghong Dai

17  Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang

18  Medical University, Urumqi, 830001, China

19  Email: epidjh@163.com

20

21

22

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

23    **Abstract**

24    **Purpose:** To investigate the potential causal link between heredity, geographical

25    environment, diet and other lifestyle factors with long-term health consequences, we

26    established the Xinjiang multi-ethnic cohort study (XMC), the first large-scale

27    prospective cohort in Xinjiang, China.

28    **Participants**: XMC commenced in 2018 and enrolled participants from three study

29    sites (Urumqi, Hotan, and Ili) in Xinjiang, China. Data collected include standard

30    baseline questionnaire, physical measurement, biological specimen. In addition, about

31    one-third of participants were assessed habitual diet by a more detailed semi-

32    quantitative food frequency questionnaire (FFQ) which included 127 foods items at

33    baseline

34    **Findings to date:** Finally, a total of 30,949 participants, with 32.37% from Urumqi,

35    41.75% from Hotan, and 25.88% from Ili were recruited in XMC. The average age of

36    participants was 56.21 years for men, and 54.75 years for women. More than 60% of

37    participants in all three survey sites reported an average consumption of fruit and

38    vegetable three or more times per week. In Hotan and Ili, the staple food was wheaten

39    food, whereas, in Urumqi, rice and wheaten food was the food staples. Consumption

40    of white meat, such as fish and poultry, was lower in the three survey locations. Based

41    on self-reported disease from study participants, the five most common chronic

42    diseases among participants across all three survey locations were dyslipidemia,

43    hypertension, cholecystitis, diabetes, ischaemic heart disease.

44 **Future plans:** We will take advantage of the existing monitoring systems or database

45 of Xinjiang, China to obtain information on health outcomes (such as morbidity and

46 death events) for all participants of the cohort study. Repeated cross-sectional surveys

47 of study participants will be conducted on a bi-annual basis with surveys focusing on

48 research outcome being the primary concern.

49 **Key words:** Cohort Profile; Dietary habit; Chronic diseases; Longitudinal Cohort

50 **Strengths and limitations of this study**

51 • The Xinjiang multi-ethnic cohort is the first population cohort study of its kind

52 established in Xinjiang, China. The study includes a study population of more

53 than 30,000 people, of which about one third are urban residents, one third are

54 farmers, and the remaining third are animal herdsmen, which is broadly

55 representative of the demographics of residents in Xinjiang.

56 • This research commenced in 2018 and participants will be surveyed once a year

57 until the end of the four-year research period (2018-2022). Data collected from

58 study participants (disease, death, migration, etc.) were verified by comparison

59 with the disease and death monitoring data network of the health and family

60 planning department, the medical record data of the hospital, the resident medical

61 insurance system, and the medical institution, as well as annual household

62 registration and death registration data of the public security department. In the

63 final year of the cohort study, 10% of subjects will be randomly selected for

64 telephone or face-to-face follow-up. After the research period ended, we will

65 keep survey the subjects by a comprehensive physical examination project within

66 Xinjiang.

67 • This study collected information on the demographic characteristics of the study

68 participants by way of a survey questionnaire, as well as blood samples. Some

69 research subjects from Ili region retained genetic material such as RNA. About

70 two-thirds of the subjects of this study are ethnic minorities. Their ethnic

71    background allows us to compare health outcomes and health determinants of

72    populations living in distinct geographic locations. The results of this study found

73    differences in the dietary habits of people from different ethnic backgrounds and

74    geographical environment which allows us to hypothesize about associations

75    between these factors and their interactions with health outcomes.

76  • The study data are not freely available, but specific proposals for future

77    collaborations are welcome. Address to the research leader of Xinjiang multi-

78    ethnic cohort study (Dai Jianghong, Email: epdjh@163.com).

79  **Introduction**

80     Over the past several decades, China has established several large-scale

81  prospective population cohort studies. These have included the China Kadoorie

82  Biobank (CKB) study[1], a population study of 500,000 people among ten provinces

83  in China, and the Shandong Multi-center Longitudinal Cohort for Health

84  Management[2], which included a research cohort of 100 million people and covered

85  a study period of 12 years. The Taizhou Longitudinal Study (TZL)[3] and Tianjin

86  Chronic Low-Grade Systemic Inflammation and Health Cohort Study (TCLSIH)[4]

87  have also been among the longitudinal cohort studies established during the same

88  period. These cohort studies have focused on analyzing the interactions between

89  health-related risk factors, as well as environmental and genetic factors, with

90  population health in China[5]. The main research outputs of these cohort studies have

91  had a substantial impact on public health promotion in China, however, to date no

92  prospective cohort studies have been carried out in the Xinjiang Uygur Autonomous

93  Region (Xinjiang) (Figure 1).

94     Xinjiang is the largest provincial administrative region in China, accounting for

95   one-sixth of China's landmass (1.66 million square kilometers). The Altai Mountains

96   in the north of Xinjiang the Kunlun Mountains in the south and the Tianshan

97   Mountains in the middle divide the region into two distinct geographic areas; the

98   Tarim Basin in the South and the Junggar Basin in the north. The topography of

99   Xinjiang is complex with geographical features varying from mountains and basins to

100  grasslands, deserts, and oases. Xinjiang has a multi-ethnic population with

101  approximately 20 million people (about 60% of the total population of Xinjiang)

102  belonging to Uyghur, Kazakh, Hui, and other non-Han ethnic groups [6] and a unique

103  diet which may be correlated with health outcomes in the region.

104      The main risk factors associated with non-communicable chronic diseases are

105  modifiable risk factors such as poor diet, lack of physical activity, and alcohol and

106  tobacco use. High-quality epidemiological data on diet, environmental and genetic

107  determinants of non-communicable chronic diseases, and long-term outcomes are

108  essential for developing public health strategies to reduce the burden of non-

109  communicable diseases. Results of statistical analyses carried out on surveillance data

110  found that heart disease, malignant tumors, and cardiovascular disease were the main

111  contributors to the overall burden of chronic health conditions in 2015, and accounted

112  for 77.67% of all deaths in Xinjiang.[7]

113      While extensive research on the etiology of chronic diseases has been carried out

114  across much of mainland China, the risk factors underpinning chronic disease may

115  vary between populations, and substantial uncertainty exists as to how important these

116  risk factors are in different settings. Because of the unique geographic environment,

117　genetic background and population make-up of Xinjiang, the etiologies of chronic

118　disease deduced from other cohort studies carried out in China may therefore not be

119　fully applicable to the population of this region. The unique geographic environment,

120　genetic background, and population make-up also lead to the special eating habits of

121　the residents in Xinjiang. The main objectives of this study were to identify diet risk

122　factors and other determinants of non-communicable diseases among the population

123　of Xinjiang.

124　**Material and Methods**

125　*Study design and Organization*

126　　　The study was conducted in collaboration with the Xinjiang Medical University,

127　the Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous

128　Region, and the Urumqi Municipal Center for Disease Control and Prevention, and

129　was completed under the guidance of Xi'an Jiaotong University. Ethical approval for

130　this study was granted by the Ethics Committee of Academy of Traditional Chinese

131　Medicine in Xinjiang Uygur Autonomous Region (2018XE0108).

132　　　The locations selected for inclusion in the cohort study, Urumqi, Hotan, and Ili,

133　were chosen by the geographical distribution of minority populations in the Xinjiang

134　Uygur Autonomous Region (Figure 2). Urumqi city is the capital of Xinjiang and is

135　the cultural and political center of the region. Surveys on factors such as health

136　conditions, health-related behaviors, and diet were carried out in 20 community health

137　service centers in Urumqi to be representative of the urban population of Xinjiang. In

138　Hotan, surveys were carried out in local villages where more than 95% of residents

139 are of the Uyghur ethnicity. The highest level of educational attainment among this

140 population is generally low and farming is the predominant occupation in this area. In

141 addition to this, the population living in Hotan has distinctive dietary habits, the health

142 outcomes of which may be possible to establish by comparing diet and health

143 outcomes between this population and populations outside of this survey location. In

144 Ili, surveys were conducted in townships where the distribution of Han, Uyghur,

145 Kazak, and Hui is relatively balanced. Most study participants recruited in this area

146 were animal herders, the survey results, therefore, being reflective of health status and

147 health-related risk factors of herdsmen in rural areas of Xinjiang. These diverse

148 survey sites were selected to be reflective of the health status of both urban and rural

149 residents, workers or farmers and herders, and different ethnicities in Xinjiang.

150 *Study participants*

151     Our survey sites were set up in community health centres or village clinics with

152 medical qualifications. For the selection of the study participants, we did not conduct

153 a randomized sampling. We conducted extensive publicity campaigns, such as

154 delivering and broadcasting introduction letters, to engage people in the study. And

155 our baseline survey was conducted in tandem with the Xinjiang Universal Health

156 Examination. Eligibility criteria for inclusion in the cohort study were adults aged 35

157 to 74 (born between 1943 and 1982), without any physical or communication

158 disabilities (Such as earning disabilities, language disorders, and muscular dystrophy),

159 with the ability to formally consent to participation in the study cohort, residing

160 permanently (at least 1 year of residency) in study sites, and whose disease incidence

161    data belonged to the local department of health. We excluded migratory populations

162    and temporary residents, as these participants are more likely to be lost to follow-up.

163    We also excluded armed forces and staff members stationed in the investigation site,

164    as long-term follow-up is also challenging among this population. All participants

165    gave written informed consent before the study began. Firstly, the investigator

166    introduced the participants to the information about XMC. Secondly, the investigator

167    asked the study participants if they would accept a questionnaire, a physical

168    examination, a biospecimen collection, and a search of their health information from a

169    medical database. Finally, the subjects were given the option to withdraw at any time

170    during the survey.

171    *Patient and public involvement*

172         Patients or the public were not involved in the design, recruitment, conduct,

173    reporting, or dissemination plans of our research. However, our researchers asked them

174    whether they would be willing to participate in this study and they were told the

175    approximate time required to participate, including length of survey response time and

176    the number of potential years of involvement.

177    *Follow-up survey*

178         Beginning in September 2016, Xinjiang's government launched a comprehensive

179    physical examination project within the region. All residents in Xinjiang can

180    participate in this free annual physical examination, with the cost covered by the local

181    government. The project will continue for decades. This free annual health

182    examination for all residents in Xinjiang included but were not limited to physical

183 examination, health-risk survey, biochemical blood examination, ultrasound, and X-

184 ray. During the past four years, Xinjiang has set up more than 2,800 physical

185 examination centres and invested a total of 3.90 billion yuan in the Xinjiang Universal

186 Health Examination. These Universal Health Examination are routinely attended by

187 more than 80% of adults residing in the region. More than 50 million people have

188 taken part in universal health examinations up to September 2018. Results of physical

189 examination have formed a complete longitudinal database. After the cohort baseline

190 survey, we can link with the database through individual identification (ID) numbers

191 to follow our participants. We have matched the baseline survey data of XMC

192 subjects with the personnel database of the national physical examination within

193 Xinjiang in 2019. More than 70% of subjects in XMC participated in the national

194 physical examination within Xinjiang in 2019.

195 Participants' follow-up will conduct by recording unique ID numbers associated

196 with everyone across several common data systems. These data systems include

197 Medical record information system, Medical insurance management system, Maternal

198 and Child health information system, Public security household registration

199 management systems, Civil affairs management system, Chronic disease management

200 information system, National central cancer registry of china, and Death registration

201 information management system. Data from these different sources can be used

202 concordantly for mutual review and for supplementing survey data.

203 In the final year of the cohort study, 10% of subjects were randomly selected for

204 telephone or face-to-face follow-up. Study participants given priority for this follow-

205 up included those that had not had any measurable outcome during the study period.

206 There are two main purposes for this strategy. Firstly, we can reduce the loss of

207 follow-up as much as possible. Secondly, we can check whether the medical database

208 can cover all participants by comparing the results of telephone or face-to-face

209 follow-up with the medical database matching.

210 *Data and blood samples collection*

211 The data collection methods applied for the Xinjiang multi-ethnic cohort study

212 included a questionnaire survey, physical examination, and collection of biological

213 samples; measures which are essential for any longitudinal study. The questionnaire

214 survey was conducted by medical students who had received consent training, and

215 physical examinations were conducted by trained nurses or doctors.

216 The baseline questionnaire we used was mainly referred to as the baseline

217 questionnaire of CKB[8] and has been slightly modified according to the opinion of

218 experts from the medical colleges in Northwest China. The questionnaire collected

219 information on sociodemographics, tea and coffee consumption, alcohol intake,

220 tobacco use, dietary habits, passive smoking and indoor air pollution, personal and

221 family medical history, physical activity, mental health, and reproductive history

222 (female study participants only; Table 1). But our collaborators conducted an

223 additional dietary survey of the study participants in the Ili region, the data collected

224 on dietary habits included specific information as to the types of food consumed, as

225 well as the frequency and intake of specific food types. For example, information on

226 the frequency and intake amount of fruit consumed, as well as specifics pertaining to

227     fruit type, i.e., number of apples, bananas, and oranges consumed, was collected.

228        Table 1. Questionnaire data collected in the Xinjiang multi-ethnic cohort study

| Questionnaire survey | Description |
|---|---|
| Demographic data | Name, Gender, ID card, Medical insurance, Date of birth, Education level, Marital status, Occupation, Financial income |
| Tea and coffee consumption | Frequency of tea drinking in the past year, Previous tea drinking habits, Types of tea drinking, Frequency of coffee consumption |
| Alcohol intake | Frequency and amount of drinking in the past year, previous drinking habits (drinking refers to drinking in the past year under normal circumstances, not holidays or special periods, such as during marriage) |
| Tobacco use | Frequency and amount of smoking in the past year, previous smoking habits (current smoker defined as daily or almost daily smoking; former smoker defined as stopping smoking for at least 6 months; non-smoker defined as that the cumulative amount of smoking in lifetime does not exceed 100) |
| Dietary habits | Intake of 30 common foods, including staple foods, animal and plant foods, soy products, dairy products, vegetables, and fruits; Types of daily consumption oil; Use of nutrient supplements; Spicy food and vinegar consumption; Household refrigerator use time |
| Passive smoking and indoor air pollution | Whether had history of live with current smoker; Frequency of weekly exposure to passive smoking and cumulative exposure time; Cooking, heating, household fuel use, and house decoration situation; Occupational exposure history of air pollutants |
| Personal and family medical history | Self-health evaluation; chest and respiratory symptoms; history of 18 common diseases including diabetes, acute myocardial infarction, hypertension, asthma, and malignant tumors; history of blood transfusion and constipation; history of 5 common chronic diseases among family members |
| Physical activity | Work-related physical activity in the past year; the way of commuting and time spent on commuting; amateur physical exercise; housework activities; frequency and length of physical activity (i.e., sweating, heartbeat, etc.); weight loss in the past year |
| Mental health | Satisfaction with current living conditions; 10 major events that may have a serious impact on life in the past two years; sleep conditions (including insomnia, daily sleep time, whether to snoring while sleeping, etc.); depression, anxiety disorders, and unexplained phobias; quality of life |
| Reproductive history (only for females) | Menstrual history of female (age at menopause and menarche), pregnancy, fertility, breastfeeding, contraceptive use, and surgery |

229        Data collected during the physical examination included height (using a medical

230     height gauge with an accuracy of 0.1cm), weight (using a medical electronic scale

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

231  with an accuracy of 0.1kg), waist circumference (using a soft measuring tape with an

232  accuracy of 0.1 cm; the lower edge of the measuring tape was placed at the highest

233  points on both sides of the hip bone, and horizontal measurement of the waist

234  circumference taken), heart rate and blood pressure (measured using a medical arm

235  electronic blood pressure monitor with an accuracy of 2 mmHg; two measurements

236  per study participant were taken, following a rest period of five to ten minutes), body

237  fat composition (measured using a body composition analyzer, TANITA DC-

238  430MA ). Participants were asked to remove jackets, shoes, and hats while all

239  physical measurements were being taken.

240      A vacuum blood collection device with intravenous anticoagulant was used to

241  collect a 20 ml blood sample for each participant. A 4 ml blood sample was used for

242  biochemical examination and routine blood examination. These examinations were

243  completed at the nearest township health service center to the survey location. Whole

244  blood samples (3 ml) were transferred to three cryopreservation tubes immediately

245  after blood samples were collected. Blood samples used to separate plasma and white

246  blood cells were centrifuged within two hours of blood sample collection (4℃ at

247  3000 rpm for 10 minutes). Samples were stored at -196℃ in liquid nitrogen

248  containers before and after bi-monthly transportation to Urumqi. RNA protection

249  solution was also added to some samples for subsequent RNA detection. All the blood

250  and urine samples are placed in a special biological sample bank, which is managed

251  by a specially assigned person using an electronic management system.

252      For a large study such as this to be practicable and economically feasible in a

253    resource-poor setting, study procedures needed to be simple and streamlined. In this

254    study, the full assessment carried out at recruitment, including obtaining informed

255    consent, administering the questionnaire survey, carrying out the physical

256    examination, and obtaining blood samples, took an average of 60 minutes to

257    complete. Investigators from all three survey locations were trained by the same

258    junior investigator and postgraduate of the Xinjiang Medical University. The survey

259    equipment was also consistent across all three survey locations and equipment

260    calibration was carried out every day.

261    *Statistical analyses*

262        For the baseline profile, descriptive statistics were calculated for baseline data

263    regarding demographic data, tea and coffee consumption, alcohol intake, tobacco use,

264    dietary habits, passive smoking and indoor air pollution, personal and family medical

265    history, physical activity, mental health, reproductive history (only for female). All

266    analyses were conducted using SAS version 9.4.

267    **Results**

268        When a participant came to the assessment center and met our eligibility criteria,

269    they were first given a serial number, and a total of 31,778 participants were given a

270    number, but some of them dropped out during the questionnaire or refused to collect a

271    biological specimen. Finally, a total of 30,949 participants were recruited as part of

272    the Xinjiang multi-ethnic cohort study from Urumqi, Hotan, and Ili in Xinjiang,

273    China. The number of participants recruited from Urumqi, Hotan, and Ili was 10017,

274    12921, and 8011, respectively (Table 2). The average age of study participants was

275   56.21 years for men, and 54.75 years for women. In Urumqi, more than half of the

276   subjects were aged over 65 years old, significantly higher than the other two regions.

277   According to the Census data of Xinjiang[9], Urumqi has the highest proportion of

278   people aged 65 and over compared to the other two regions. Also, our baseline survey

279   was conducted on a weekday, but in urban areas, younger people are more likely to be

280   required to work on a weekday. The proportion of Han Chinese in the Urumqi study

281   population was close to 90%, while the proportion of Uyghur in the Hotan study

282   population was over 99%. The proportion of ethnic groups represented in the study

283   population recruited in the Ili region, conversely, was relatively balanced, indicating

284   that ethnic composition varies significantly from region to region. The average level

285   of highest educational attainment and average income among the Urumqi population

286   was significantly higher than that of the Hotan and Ili populations. The proportion of

287   people who had been educated to primary or pre-primary level in Hotan and Yili were

288   87.65% and 71.14%, respectively.

289 Table 2. Demographic characteristics of the study participants by site

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Men** | **Women** | **Total** | **Men** | **Women** | **Total** | **Men** | **Women** | **Total** |
| Age(years) | | | | | | | | | |
| <45 | 231(5.87) | 330(5.43) | 561(5.60) | 1040(21.17) | 2337(29.18) | 3377(26.14) | 1120(29.60) | 1250(29.57) | 2370(29.58) |
| 45-54 | 570(14.47) | 1021(16.80) | 1591(15.88) | 1408(28.66) | 2871(35.85) | 4279(33.12) | 1248(32.98) | 1537(36.36) | 2785(34.76) |
| 55-64 | 824(20.92) | 1358(22.34) | 2182(21.78) | 1364(27.76) | 1848(23.08) | 3212(24.86) | 870(22.99) | 994(23.52) | 1864(23.27) |
| 65- | 2313(58.74) | 3370(55.44) | 5683(56.73) | 1101(22.41) | 952(11.89) | 2053(15.89) | 546(14.43) | 446(10.55) | 992(12.38) |
| Ethnicity | | | | | | | | | |
| Han | 3457(87.79) | 5432(89.36) | 8889(88.74) | 6(0.12) | 6(0.07) | 12(0.09) | 572(15.12) | 522(12.35) | 1094(13.66) |
| Hui | 310(7.87) | 392(6.45) | 702(7.01) | 4(0.08) | 15(0.19) | 19(0.15) | 1336(35.31) | 1333(31.54) | 2669(33.32) |
| Uyghur | 105(2.67) | 168(2.76) | 273(2.73) | 4897(99.67) | 7969(99.51) | 12866(99.57) | 908(24.00) | 1361(32.20) | 2269(28.32) |
| Kazakh | 41(1.04) | 53(0.87) | 94(0.94) | 6(0.12) | 13(0.16) | 19(0.15) | 883(23.34) | 951(22.50) | 1834(22.89) |
| Other | 25(0.63) | 34(0.56) | 59(0.59) | NA | 5(0.06) | 5(0.04) | 85(2.25) | 60(1.42) | 145(1.81) |
| Education | | | | | | | | | |
| Primary/less | 1064(27.02) | 2534(41.68) | 3598(35.92) | 4093(83.31) | 7232(90.31) | 11325(87.65) | 2582(68.23) | 3117(73.74) | 5699(71.14) |
| Secondary | 2320(58.91) | 3020(49.68) | 5340(53.31) | 774(15.75) | 734(9.17) | 1508(11.67) | 1166(30.81) | 1071(25.34) | 2237(27.92) |
| Degree or above | 553(14.04) | 517(8.50) | 1070(10.68) | 32(0.65) | 18(0.22) | 50(0.39) | 28(0.74) | 26(0.62) | 54(0.67) |
| Refused/missing | 1(0.03) | 8(0.13) | 9(0.09) | 14(0.28) | 24(0.30) | 38(0.29) | 8(0.21) | 13(0.31) | 21(0.26) |

15

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Men | Women | Total | Men | Women | Total | Men | Women | Total |
| Marriage | | | | | | | | | |
| Married | 3696(93.85) | 5045(82.99) | 8741(87.26) | 4319(87.91) | 6165(76.99) | 10484(81.14) | 3500(92.49) | 3586(84.84) | 7086(88.45) |
| Widowed/separated | 197(5.00) | 1006(16.55) | 1203(12.01) | 538(10.95) | 1773(22.14) | 2311(17.89) | 227(6.00) | 614(14.53) | 841(10.50) |
| Never married | 42(1.07) | 19(0.31) | 61(0.61) | 19(0.39) | 6(0.07) | 25(0.19) | 43(1.14) | 6(0.14) | 49(0.61) |
| Refused/missing | 3(0.08) | 9(0.15) | 12(0.12) | 37(0.75) | 64(0.80) | 101(0.78) | 14(0.37) | 21(0.50) | 35(0.44) |
| Income (RMB)* | | | | | | | | | |
| under 10000 | 632(16.05) | 912(15.00) | 1544(15.41) | 2689(54.73) | 5139(64.17) | 7828(60.58) | 414(10.94) | 616(14.57) | 1030(12.86) |
| 10000-19999 | 412(10.46) | 585(9.62) | 997(9.95) | 1196(24.34) | 1649(20.59) | 2845(22.02) | 1311(34.65) | 1712(40.50) | 3023(37.74) |
| 20000-34999 | 411(10.44) | 737(12.12) | 1148(11.46) | 453(9.22) | 501(6.26) | 954(7.38) | 1157(30.58) | 1093(25.86) | 2250(28.09) |
| 35000-49999 | 504(12.80) | 885(14.56) | 1389(13.87) | 148(3.01) | 198(2.47) | 346(2.68) | 393(10.39) | 339(8.02) | 732(9.14) |
| 50000 or above | 1978(50.23) | 2952(48.56) | 4930(49.22) | 416(8.47) | 500(6.24) | 916(7.09) | 503(13.29) | 459(10.86) | 962(12.01) |
| Refused/missing | 1(0.03) | 8(0.13) | 9(0.09) | 11(0.22) | 21(0.26) | 32(0.25) | 6(0.16) | 8(0.19) | 14(0.17) |
| Anthropometric index | | | | | | | | | |
| Height | 169.03±6.26 | 157.98±19.33 | 162.33±16.47 | 163.32±7.29 | 153.60±6.81 | 157.30±8.44 | 167.13±6.98 | 155.98±7.05 | 161.24±8.95 |
| Weight | 73.82±10.69 | 63.10±11.46 | 67.31±12.33 | 66.71±12.38 | 59.63±12.04 | 62.32±12.64 | 72.24±12.06 | 65.20±12.33 | 68.52±12.69 |
| Waist circumference | 92.40±17.00 | 87.09±15.47 | 89.18±16.30 | 91.88±11.50 | 91.44±12.29 | 91.61±12.00 | 90.56±11.02 | 88.49±11.80 | 89.46±11.48 |

16

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Men | Women | Total | Men | Women | Total | Men | Women | Total |
| Body mass index | 25.81±3.31 | 25.35±4.42 | 25.53±4.02 | 24.96±4.04 | 25.24±4.67 | 25.14±4.44 | 25.82±3.75 | 26.76±4.51 | 26.31±4.20 |

290    N(%) for categorical variable; Mean ± SD for continuous variables; *: 1 RMB = 0.14524 USD and 1 RMB = 0.1277 EUR in December 2018.

17

291          Approximately 50% of participants surveyed in Urumqi had an average annual

292     income of more than 50,000, whereas about 60% of subjects in Hotan had an average

293     annual income of less than 10,000. Study participants in Urumqi were found to have

294     the highest average height, while study participants in Ili were found to have the

295     highest average weight, with study participants in Hotan having the highest average

296     waist circumference. The average body mass index of the subjects in the three regions

297     exceeded the diagnostic threshold (BMI $\geq$ 25kg/m$^2$)[10] for overweight in China,

298     with study participants in the Ili region having the highest average body mass index.

299          The five most common chronic diseases among participants across all three

300     survey sites were dyslipidemia, hypertension, cholecystitis, diabetes, ischaemic heart

301     disease with a prevalence of 34.55%, 32.66%, 14.34%, 10.07%, and 8.74%,

302     respectively (Table 3).

303          Table 3. Self-reported disease prevalence among subjects in different sites (%)

| Diseases | Urumqi | | Hotan | | Ili | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Crude | Adjusted | Crude | Adjusted | Crude | Adjusted | Crude | Adjusted |
| Cardiovascular diseases | | | | | | | | |
| Hypertension* | 47.19 | 29.29 | 37.00 | 32.64 | 38.61 | 36.09 | 40.74 | 32.66 |
| Hypertension | 32.64 | 17.87 | 27.35 | 23.76 | 16.71 | 15.23 | 26.31 | 19.80 |
| IHD | 8.18 | 3.28 | 13.54 | 12.15 | 8.44 | 7.72 | 10.48 | 8.74 |
| Other heart disease | 0.57 | 0.32 | 2.63 | 2.34 | 0.69 | 0.62 | 1.46 | 1.27 |
| Stroke | 3.43 | 1.62 | 4.91 | 4.31 | 1.61 | 1.47 | 3.58 | 2.82 |
| Metabolic disease and kidney diseases | | | | | | | | |
| Diabetes* | 23.93 | 19.14 | 7.26 | 6.74 | 9.80 | 9.19 | 13.11 | 10.07 |
| Diabetes | 14.73 | 8.81 | 4.25 | 3.87 | 4.69 | 4.36 | 7.76 | 5.54 |
| Dyslipidemia* | 35.00 | 35.08 | 33.53 | 34.05 | 34.83 | 35.08 | 34.32 | 34.55 |
| CKD | 0.68 | 0.56 | 6.20 | 6.28 | 0.99 | 0.93 | 3.06 | 3.17 |
| Anemia* | 3.31 | 5.30 | 5.89 | 5.55 | 5.30 | 5.26 | 4.98 | 5.44 |
| Chronic respiratory diseases | | | | | | | | |
| Chronic bronchitis | 4.48 | 2.16 | 15.89 | 14.17 | 3.31 | 3.12 | 8.94 | 7.87 |
| Emphysema | 0.29 | 0.15 | 3.00 | 2.57 | 0.51 | 0.49 | 1.48 | 1.25 |
| COPD | 0.40 | 0.17 | 1.22 | 1.14 | 0.20 | 0.18 | 0.69 | 0.62 |
| Asthma | 1.11 | 0.65 | 2.86 | 2.45 | 0.22 | 0.20 | 1.61 | 1.36 |

18

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tuberculosis | 0.69 | 0.57 | 2.84 | 2.55 | 0.54 | 0.50 | 1.55 | 1.37 |
| Digestive diseases | | | | | | | | |
| Chronic hepatitis | 0.49 | 0.53 | 4.85 | 4.91 | 1.34 | 1.33 | 2.53 | 2.82 |
| Cholecystitis | 6.50 | 3.12 | 23.28 | 21.28 | 13.66 | 12.49 | 15.36 | 14.34 |
| Peptic ulcer | 0.91 | 0.74 | 6.61 | 6.33 | 2.11 | 2.01 | 3.60 | 3.61 |
| Skeletal disorders | | | | | | | | |
| Osteoporosis | 2.15 | 0.83 | 7.48 | 6.50 | 3.74 | 3.52 | 4.79 | 4.21 |
| Fracture | 1.93 | 1.24 | 5.82 | 6.10 | 5.75 | 5.87 | 4.54 | 4.96 |
| Cancer | 0.63 | 0.32 | 0.67 | 0.54 | 0.46 | 0.43 | 0.60 | 0.48 |

304    *Self-reported combined with biological detection

305    Adjusted: Age and gender adjusted based on 2010 China census data

306    IHD: Ischaemic heart disease, Other heart disease: included Rheumatic heart disease and Cor pulmonale,

307    CKD: Chronic kidney disease

308        Disease prevalence varies substantially across the three survey locations, which

309    may be attributed to differences in nationality, diet, education, socioeconomics, and

310    living habits between the three areas. After standardizing for age and sex based on

311    2010 China census data, the prevalence of hypertension in Ili was 36.09%, higher than

312    that in Urumqi (29.29%) and Hotan (32.64%). According to a survey carried out

313    between 2012 to 2015 the average prevalence of hypertension in China was 23%[11].

314    The prevalence of hypertension in the XMC study sites, as determined by this study,

315    was much higher than the national average. Study participants in Urumqi had a higher

316    prevalence of diabetes (19.14%) compared with those in Ili and Hotan. The results of

317    this study also found that the prevalence of diabetes in Ili and Hotan was slightly

318    lower than the national average (10.9%).[12] Urumqi has a high prevalence of

319    diabetes and hypertension, which may be associated with participants residing in

320    urban communities and having a higher average income compared with rural dwelling

321    populations in Hotan and Ili. As a result, Urumqi's people have better health resources

322    so people with diabetes are more likely to be diagnosed correctly. In addition,

323    Urumqi's people have a lower proportion of the physical labor workforce than the

19

BMJ Open

324 other two regions, so that they may be sedentary in work time, which may lead to a

325 higher prevalence of diabetes in the Urumqi population. The prevalence of chronic

326 kidney disease among participants in Hotan (6.28%) was approximately ten times

327 higher than that of Urumqi (0.56%) and six times that of Ili (0.93%). The prevalence

328 of cholecystitis in Hotan (21.28%) was approximately seven times that of the

329 prevalence in Urumqi (3.12%). Results of this study also found that the prevalence of

330 COPD in Xinjiang (0.62%) was much lower than the national level (3.84%)[13].

331 Regarding Cholecystitis, the data on the national average was limited. A previous

332 epidemiological survey showed that the incidence of gallstones in Xinjiang was

333 11.83%.[14] The prevalence of Cholecystitis in Ili (12.49%) and Hotan (21.28%) was

334 much higher than previously recorded levels. This may be associated with Hotan's

335 local dietary habits. The prevalence of ischemic heart disease, stroke, chronic

336 hepatitis, and CKD in Hotan was also found to be higher than that of the other two

337 sites.

338 About baseline dietary intake (see Figure 3), more than 60% of participants in all

339 three survey locations reported an average consumption of fruit and vegetable three or

340 more times per week. In Hotan and Ili, the staple food was wheaten food, whereas, in

341 Urumqi, rice and wheaten food was the food staples. Dietary habits in Hotan, which

342 has a predominantly Uyghur population, consisted predominantly of mutton, with

343 more than 70% consuming mutton three or more times per week. The proportion of

344 the population who consumed pork, mutton, or beef three or more times per week in

345 Urumqi by contrast, where the population is predominantly Han, was 28.26%, 20.24,

20

346  and 24.28%, respectively. Consumption of white meat, such as fish and poultry, was

347  lower in the three survey locations. Fresh meat consumed by Xinjiang residents

348  consisted mainly of red meat such as beef and mutton. Some studies have suggested

349  that too high an intake of red meat might increase the risk of metabolic syndrome

350  [15,16]. The study also showed that consumption of fresh milk was associated with a

351  lower risk of Metabolic Syndrome[16]. In this study, the proportion of people

352  drinking milk three or more times per week was lower in Ili and particularly in Hotan

353  (<10% of participants surveyed). More than 60% of people in Ili reported drinking

354  milk tea ≥3 times a week, however, with few of them drinking fresh milk directly.

355       A diet high in salt and fat is associated with an increased risk of high blood

356  pressure and ischemic heart disease[17]. The consumption of coarse grains, eggs, milk

357  and yogurt three or more times per week among people surveyed in Urumqi was

358  higher than among people surveyed in Ili and Hotan. In addition to this, consumption

359  of fowl, fish or seafood, and beans three or more times per week was low in all three

360  survey locations (<10% of participants surveyed). The results of the survey on dietary

361  habits reported here are generally consistent with a recent survey[18], which found

362  that consumption of cereals, meat, and oil in Xinjiang exceeded maximum

363  recommended allowances, while consumption of eggs, seafood products, dairy

364  products, nuts, and legumes were below the minimum recommended allowances.

365  Notably, this study reported higher consumption of fruit and vegetable in all three

366  regions than previous studies on fruit and vegetable consumption in Xinjiang[16,19].

367  **Discussion**

21

368        This is the largest prospective study that has been undertaken to date in Xinjiang,

369 China. The main strength of the Xinjiang multi-ethnic cohort study is the inclusion of

370 over 20,000 survey participants from diverse ethnic backgrounds such as Uyghur,

371 Kazakh, and Hui, which allows prevalence of chronic conditions, as well as dietary

372 habits, various health determinants, and socio-demographics between distinct

373 populations to be compared. It is already well established that the dietary habits of

374 Uyghur, Kazakh, and Hui differ from those of Han Chinese. We have investigated the

375 dietary habits of all the study participants as part of this cohort study, in conjunction

376 with prevalence of non-communicable diseases to compare health determinants and

377 prevalence of chronic conditions between these three survey locations.

378        Another important strength of this study was the survey questionnaire, which

379 collected details on major risk factors such as alcohol intake, tobacco use (we

380 investigated current as well as past usage quantity, frequency, type), mental health,

381 and quality of life. The three survey locations selected for this study were broadly

382 representative of the different populations residing in Xinjiang, including animal

383 herders, and urban and rural residents. The results of the survey conducted here can be

384 used to compare differences in prevalence and type of non-communicable chronic

385 diseases among these three populations.

386        There were also several limitations with the Xinjiang multi-ethnic cohort study.

387 Firstly, we did not use random sampling to select study participants, but rather a

388 convenience sampling strategy whereby eligible study participants from

389 predetermined survey locations were recruited until an adequate sample size had been

22

390   met. We excluded those who had lived in the local area for less than one year, which

391   may reduce how representative the study population is for Urumqi regions. However,

392   for the Ili and Hotan regions, the population is more stable and may not have an

393   impact on the representation of the population. As this study is the first large-scale

394   cohort study to have been carried out in Xinjiang, and data collected as part of

395   previous cross-sectional health surveys of Xinjiang residents have not yet been

396   published, we cannot compare differences in the distribution of diseases and related

397   factors between our cohort study population and the wider Xinjiang population.

398   Another limitation is that diseases and exposures were self-reported. Although a

399   detailed and comprehensive questionnaire was carried out, some results are subject to

400   reporting and recall bias.

401       We will take advantage of the existing monitoring systems or database of

402   Xinjiang, China to obtain information on health outcomes (such as morbidity and

403   death events) for all participants of the cohort study. Repeated cross-sectional surveys

404   of study participants will be conducted on a bi-annual basis with surveys focusing on

405   research outcome being the primary concern. Data will be obtained by administering

406   questionnaires used for baseline research, as well as additional information on health

407   determinants. Although the study population is relatively large, the availability of

408   repeated measures, extensive biobank blood, and urine samples, and health

409   information systems linked by an ID number will allow this study to assess within-

410   and between-person variability over time in major risk factors among the population

411   of people living in Xinjiang, China.

23

412      Baseline data from the Xinjiang multi-ethnic cohort study are put through a

413 process of editing. The study data are not freely available, but specific proposals for

414 future collaborations are welcome. Address to the research leader of Xinjiang multi-

415 ethnic cohort study.

422 **Contributorship statement：** Conceptualization: Hong Yan, Jianghong Dai;

423 Methodology: Jianghong Dai, Hong Yan, Qi Sun, Gaofeng Sun; Software: Tao Luo;

424 Validation: Tian Tian; Formal analysis: Tao Luo, Tian Tian; Investigation: Tao Luo,

425 Zewen Zhang, Liu, Lirong, Qi Sun, Gaofeng Sun; Resources: Jianghong Dai, Hong

426 Yan; Data curation: Tao Luo, Zewen Zhang, Liu, Lirong; Writing-original draft

427 preparation: Tao Luo, Tian Tian, Zewen Zhang, Liu, Lirong; Writing-review and

428 editing: Jianghong Dai; Visualization: Tao Luo; Supervision: Jianghong Dai; Project

429 administration: Jianghong Dai, Hong Yan; Funding acquisition: Jianghong Dai, Hong

430 Yan.

431 **References**

432 1   Chen Z, Chen J, Collins R, *et al.* China Kadoorie Biobank of 0.5 million people: survey methods,
433     baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;**40**:1652–66.
434     doi:10.1093/ije/dyr120

435 2   Zhang T, Zhang C, Zhang Y, *et al.* Metabolic syndrome and its components as predictors of

24

436    nonalcoholic fatty liver disease in a northern urban Han Chinese population: a prospective
437    cohort study. *Atherosclerosis* 2015;**240**:144–8. doi:10.1016/j.atherosclerosis.2015.02.049

438    3    Wang X, Lu M, Qian J, *et al.* Rationales, design and recruitment of the Taizhou Longitudinal
439        Study. *BMC Public Health* 2009;**9**:223. doi:10.1186/1471-2458-9-223

440    4    Song K, Du H, Zhang Q, *et al.* Serum immunoglobulin M concentration is positively related to
441        metabolic syndrome in an adult population: Tianjin Chronic Low-Grade Systemic
442        Inflammation and Health (TCLSIH) Cohort Study. *PLoS ONE* 2014;**9**:e88701.
443        doi:10.1371/journal.pone.0088701

444    5    Li L, Lyu J. Large prospective cohort studies: a review and update. *Zhonghua Liu Xing Bing*
445        *Xue Za Zhi* 2015;**36**:1187–9.

446    6    He J, Guo S, Liu J, *et al.* Ethnic differences in prevalence of general obesity and abdominal
447        obesity among low-income rural Kazakh and Uyghur adults in far western China and
448        implications    in    preventive    public    health.    *PLoS    ONE*    2014;**9**:e106723.
449        doi:10.1371/journal.pone.0106723

450    7    ZHE W, PUERHATI W, LIU L, *et al.* Analysis of the death causes of residents in 15
451        surveillance sites in Xinjiang in 2015. *Modern Preventive Medicine* 2017;**44**:4186–90.

452    8    Li L, Lv J, Guo Y, *et al.* The China Kadoorie Biobank: related methodology and baseline
453        characteristics of the participants. *Zhonghua Liu Xing Bing Xue Za Zhi* 2012;**33**:249–55.

454    9    Office of the Leading Group of the Xinjiang Uygur Autonomous Region for the Seventh
455        National Population Census. Communiqué of the Seventh National Population Census of the
456        Xinjiang Uygur Autonomous Region. Statistic Bureau of Xinjiang Uygur Autonomous Region
457        2021.

458    10   Reynolds K, Gu D, Whelton PK, *et al.* Prevalence and risk factors of overweight and obesity
459        in China. *Obesity (Silver Spring)* 2007;**15**:10–8. doi:10.1038/oby.2007.527

460    11   Wang Z, Chen Z, Zhang L, *et al.* Status of Hypertension in China: Results From the China
461        Hypertension        Survey,        2012-2015.        *Circulation*        2018;**137**:2344–56.
462        doi:10.1161/CIRCULATIONAHA.117.032380

463    12   Wang L, Gao P, Zhang M, *et al.* Prevalence and Ethnic Pattern of Diabetes and Prediabetes in
464        China in 2013. *JAMA* 2017;**317**:2515–23. doi:10.1001/jama.2017.7596

465    13   Chan KY, Li X, Chen W, *et al.* Prevalence of chronic obstructive pulmonary disease (COPD)
466        in China in 1990 and 2010. *J Glob Health* 2017;**7**:020704. doi:10.7189/jogh.07.020704

467    14   Zhu L, Aili A, Zhang C, *et al.* Prevalence of and risk factors for gallstones in Uighur and Han
468        Chinese. *World J Gastroenterol* 2014;**20**:14942–9. doi:10.3748/wjg.v20.i40.14942

469    15   de Oliveira Otto MC, Alonso A, Lee D-H, *et al.* Dietary intakes of zinc and heme iron from red

25

470    meat, but not from other sources, are associated with greater risk of metabolic syndrome and
471    cardiovascular disease. *J Nutr* 2012;**142**:526–33. doi:10.3945/jn.111.149781

472  16  Guo H, Gao X, Ma R, *et al.* Prevalence of Metabolic Syndrome and its Associated Factors
473      among Multi-ethnic Adults in Rural Areas in Xinjiang, China. *Sci Rep* 2017;**7**:17643.
474      doi:10.1038/s41598-017-17870-5

475  17  Zhai FY, Du SF, Wang ZH, *et al.* Dynamics of the Chinese diet and the role of urbanicity,
476      1991-2011. *Obes Rev* 2014;**15 Suppl 1**:16–26. doi:10.1111/obr.12124

477  18  Yin J, Yang D, Zhang X, *et al.* Diet shift: Considering environment, health and food culture.
478      *Sci Total Environ* 2020;**719**:137484. doi:10.1016/j.scitotenv.2020.137484

479  19  Zhai F, He Y, Wang Z, *et al.* Status and characteristic of dietary intake of 12 minority
480      nationalities in China. *Wei Sheng Yan Jiu* 2007;**36**:539–41.

481    Figure legends
482    Table 1. Questionnaire data collected in the Xinjiang multi-ethnic cohort study
483    Table 2. Demographic characteristics of the study participants by site
484    Table 3. Self-reported disease prevalence among subjects in different sites
485    Figure 1. Locations of on-going prospective cohorts in China
486    CKB: China Kadoorie Biobank; SMLC: Shandong Multi-center Longitudinal Cohort; XMC: Xinjiang Multi-ethnic Cohort; TZL: Taizhou
487    Longitudinal Study; TCLSIH: Tianjin Chronic Low-Grade Systemic Inflammation and Health Cohort Study
488    Figure 2. Map of Xinjiang, China, showing the locations of three surveyed sites (Red five-pointed star)
489    in the XMC
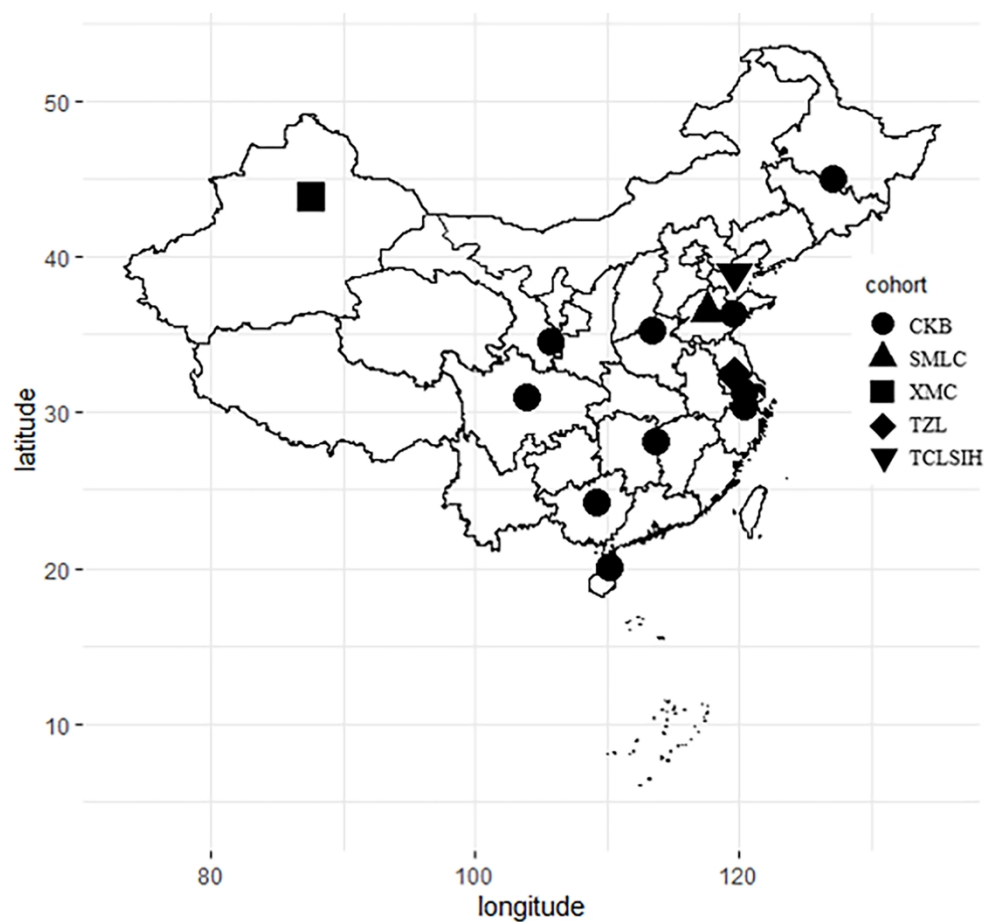490    Figure 3. Percentage of common food intake in different sites in the XMC
491

26

Figure 1. Locations of on-going prospective cohorts in China.CKB: China Kadoorie Biobank; SMLC: Shandong Multi-center Longitudinal Cohort; XMC: Xinjiang Multi-ethnic Cohort; TZL: Taizhou Longitudinal Study; TCLSIH: Tianjin Chronic Low-Grade Systemic Inflammation and Health Cohort Study
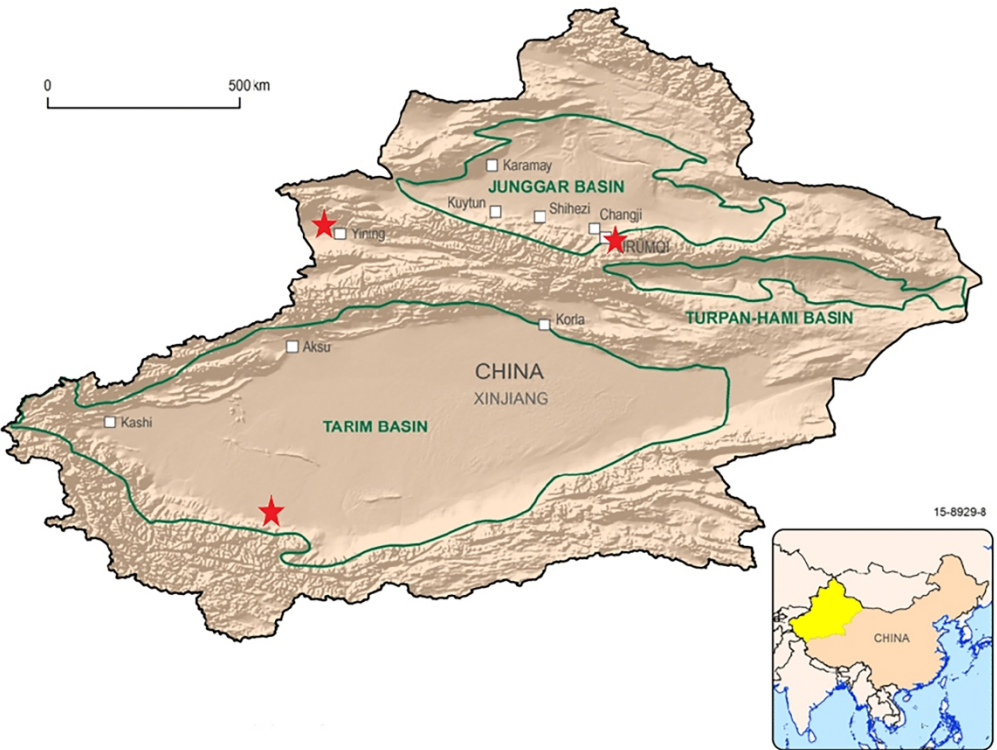
296x274mm (300 x 300 DPI)

Figure 2. Map of Xinjiang, China, showing the locations of three surveyed sites (Red five-pointed star) in the XMC
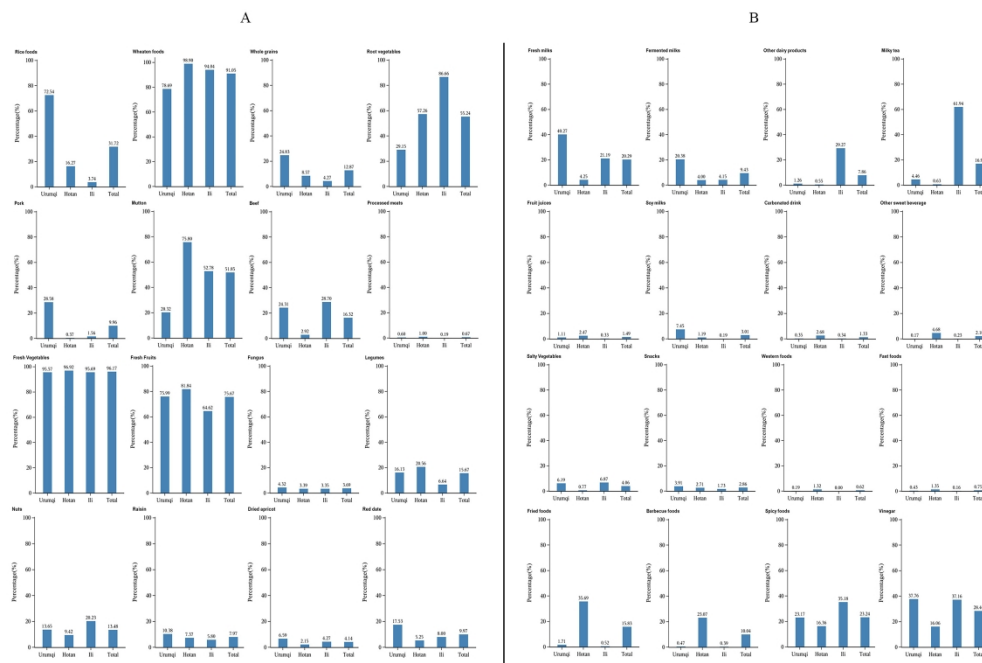
571x430mm (300 x 300 DPI)

Figure 3. Percentage of common food intake in different sites in the XMC

635x425mm (300 x 300 DPI)

# Cohort profile: The Xinjiang multi-ethnic cohort(XMC) study

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**BMJ**

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

1  **Cohort profile: The Xinjiang multi-ethnic cohort(XMC) study**

2  **Short Title:** Profile of XMC Study

3  Luo Tao[1#], Tian Tian[2#], Liu Lirong[1], Zhang Zewen[1], Sun Qi[3], Sun Gaofeng[4], Jianghong

4  Dai [1*], Yan Hong[5]

5  [1]Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang

6  Medical University, Urumqi, 830001, China;

7  [2]Shenzhen Campus of Sun Yat-sen University, School of Public Health (Shenzhen),

8  Sun Yat-sen University, Shenzhen, Guangdong, 518107, China;

9  [3]Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous Region,

10  Urumqi, 830099, China;

11  [4]Urumqi Municipal Center for Disease Control and Prevention, Urumqi, 830000, China;

12  [5]Department of Epidemiology and Health Statistics, School of Public Health of Xi'an

13  Jiaotong University Health Science Center, Xi'an, 710061, China;

14  [#]These authors contributed equally

15  *Correspondence:

16  Jianghong Dai

17  Department of Epidemiology and Biostatistics, School of Public Health, Xinjiang

18  Medical University, Urumqi, 830001, China

19  Email: epidjh@163.com

20

21

22

**Abstract**

**Purpose:** To investigate the potential causal link between heredity, geographical environment, diet and other lifestyle factors with long-term health consequences, we established the Xinjiang multi-ethnic cohort study (XMC), the first large-scale prospective cohort in Xinjiang, China.

**Participants**: XMC commenced in 2018 and enrolled participants from three study sites (Urumqi, Hotan, and Ili) in Xinjiang, China. Data collected include standard baseline questionnaire, physical measurement, biological specimen. In addition, about one-third of participants were assessed habitual diet by a more detailed semi-quantitative food frequency questionnaire (FFQ) which included 127 foods items at baseline

**Findings to date:** Finally, a total of 30,949 participants, with 32.37% from Urumqi, 41.75% from Hotan, and 25.88% from Ili were recruited in XMC. The average age of participants was 56.21 years for men, and 54.75 years for women. More than 60% of participants in all three survey sites reported an average consumption of fruit and vegetable three or more times per week. In Hotan and Ili, the staple food was wheaten food, whereas, in Urumqi, rice and wheaten food was the food staples. Consumption of white meat, such as fish and poultry, was lower in the three survey locations. Based on self-reported disease from study participants, the five most common chronic diseases among participants across all three survey locations were dyslipidemia, hypertension, cholecystitis, diabetes, ischaemic heart disease.

44     **Future plans:** Firstly, we will collect all health-related records of the study

45     participants in January each year for the previous year. Secondly, 10% of subjects

46     were randomly selected for telephone follow-up in the final year of cohort building.

47     Finally, as planned, we will revisit the study subjects on site every 2-3 years. Again,

48     we will conduct face-to-face questionnaires and collect biological specimens such as

49     blood and urine from the study subjects.

50     **Key words:** Cohort Profile; Dietary habit; Chronic diseases; Longitudinal Cohort

51     **Strengths and limitations of this study**

52     • The Xinjiang multi-ethnic cohort is the first population cohort study established

53       in Xinjiang, China, broadly representative of the demographics of residents in

54       Xinjiang.

55     • The study collected information through face-to-face questionnaires and

56       biological specimens such as blood and urine from the study subjects.

57     • We have established a high-level biospecimen bank to manage our baseline

58       samples, which will benefit our subsequent studies.

59     • We have a comprehensive follow-up schedule that includes annual medical

60       record follow-up and face-to-face visits approximately once every 2 to 3 years.

61     • The study data are not freely available, but specific proposals for future

62       collaborations are welcome. Address to the research leader of Xinjiang multi-

63       ethnic cohort study (Dai Jianghong, Email: epdjh@163.com).

64

65

66

67

**Introduction**

68

69     Over the past several decades, China has established several large-scale

70 prospective population cohort studies. These have included the China Kadoorie

71 Biobank (CKB) study[1], a population study of 500,000 people among ten provinces

72 in China, and the Shandong Multi-center Longitudinal Cohort for Health

73 Management[2], which included a research cohort of 100 million people and covered

74 a study period of 12 years. The Taizhou Longitudinal Study (TZL)[3] and Tianjin

75 Chronic Low-Grade Systemic Inflammation and Health Cohort Study (TCLSIH)[4]

76 have also been among the longitudinal cohort studies established during the same

77 period. These cohort studies have focused on analyzing the interactions between

78 health-related risk factors and environmental and genetic factors with population

79 health in China[5]. The main research outputs of these cohort studies have had a

80 substantial impact on public health promotion in China. However, no prospective

81 cohort studies have been carried out in the Xinjiang Uygur Autonomous Region

82 (Xinjiang) (Figure 1).

83     Xinjiang is China's largest provincial administrative region, accounting for one-

84 sixth of China's landmass (1.66 million square kilometres). The Altai Mountains in

85 the north of Xinjiang, the Kunlun Mountains in the south and the Tianshan Mountains

86 in the middle divide the region into two distinct geographic areas; the Tarim Basin in

87 the South and the Junggar Basin in the north. The topography of Xinjiang is complex,

88 with geographical features varying from mountains and basins to grasslands, deserts,

89 and oases. Xinjiang has a multi-ethnic population with approximately 20 million

90  people. About 60% of the total population of Xinjiang belongs to Uyghur, Kazakh,

91  Hui, and other non-Han ethnic groups [6], and a special diet may be correlated with

92  health outcomes in the region.

93      The main risk factors associated with non-communicable chronic diseases are

94  modifiable risk factors such as poor diet, lack of physical activity, and alcohol and

95  tobacco use. High-quality epidemiological data on diet, environmental and genetic

96  determinants of non-communicable chronic diseases, and long-term outcomes are

97  essential for developing public health strategies to reduce the burden of non-

98  communicable diseases. Results of statistical analyses carried out on surveillance data

99  found that heart disease, malignant tumours, and cardiovascular disease were the main

100  contributors to the overall burden of chronic health conditions in 2015, and accounted

101  for 77.67% of all deaths in Xinjiang.[7]

102      While extensive research on the aetiology of chronic diseases has been carried

103  out across much of mainland China, the risk factors underpinning chronic disease may

104  vary between populations. Substantial uncertainty exists as to how important these

105  risk factors are in different settings. Because of the unique geographic environment,

106  genetic background, and population makeup of Xinjiang, the etiologies of chronic

107  disease deduced from other cohort studies carried out in China may not be fully

108  applicable to the population of this region. The unique geographic environment,

109  genetic background, and population makeup also lead to the special eating habits of

110  the residents in Xinjiang. The main objectives of this study were to identify diet risk

111  factors and other determinants of non-communicable diseases among the population

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

112    of Xinjiang.

113    **Material and Methods**

114    *Study design and Organization*

115       The study was conducted in collaboration with the Xinjiang Medical University,

116    the Academy of Traditional Chinese Medicine in Xinjiang Uygur Autonomous

117    Region, and the Urumqi Municipal Center for Disease Control and Prevention, and

118    was completed under the guidance of Xi'an Jiaotong University. Ethical approval for

119    this study was granted by the Ethics Committee of Academy of Traditional Chinese

120    Medicine in Xinjiang Uygur Autonomous Region (2018XE0108).

121       The locations selected for inclusion in the cohort study, Urumqi, Hotan, and Ili,

122    were chosen by the geographical distribution of minority populations in the Xinjiang

123    Uygur Autonomous Region (Figure 2). Urumqi city is the capital of Xinjiang and is

124    the cultural and political center of the region. Surveys on factors such as health

125    conditions, health-related behaviors, and diet were carried out in 20 community health

126    service centers in Urumqi to be representative of the urban population of Xinjiang. In

127    Hotan, surveys were carried out in local villages where more than 95% of residents

128    are of the Uyghur ethnicity. The highest level of educational attainment among this

129    population is generally low and farming is the predominant occupation in this area. In

130    addition to this, the population living in Hotan has distinctive dietary habits, the health

131    outcomes of which may be possible to establish by comparing diet and health

132    outcomes between this population and populations outside of this survey location. In

133    Ili, surveys were conducted in townships where the distribution of Han, Uyghur,

134  Kazak, and Hui is relatively balanced. Most study participants recruited in this area

135  were animal herders, the survey results, therefore, being reflective of health status and

136  health-related risk factors of herdsmen in rural areas of Xinjiang. These diverse

137  survey sites were selected to be reflective of the health status of both urban and rural

138  residents, workers or farmers and herders, and different ethnicities in Xinjiang.

139  *Study participants*

140      Our survey sites were set up in community health centres or village clinics with

141  medical qualifications. For the selection of the study participants, we did not conduct

142  a randomized sampling. We conducted extensive publicity campaigns, such as

143  delivering and broadcasting introduction letters, to engage people in the study. And

144  our baseline survey was conducted in tandem with the Xinjiang Universal Health

145  Examination. Eligibility criteria for inclusion in the cohort study were adults aged 35

146  to 74 (born between 1943 and 1982), without any physical or communication

147  disabilities (Such as earning disabilities, language disorders, and muscular dystrophy),

148  with the ability to formally consent to participation in the study cohort, residing

149  permanently (at least 1 year of residency) in study sites, and whose disease incidence

150  data belonged to the local department of health. We excluded migratory populations

151  and temporary residents, as these participants are more likely to be lost to follow-up.

152  We also excluded armed forces and staff members stationed in the investigation site,

153  as long-term follow-up is also challenging among this population. All participants

154  gave written informed consent before the study began. Firstly, the investigator

155  introduced the participants to the information about XMC. Secondly, the investigator

156 asked the study participants if they would accept a questionnaire, a physical

157 examination, a biospecimen collection, and a search of their health information from a

158 medical database. Finally, the subjects were given the option to withdraw at any time

159 during the survey.

160 *Patient and public involvement*

161     Patients or the public were not involved in the design, recruitment, conduct,

162 reporting, or dissemination plans of our research. However, our researchers asked them

163 whether they would be willing to participate in this study and they were told the

164 approximate time required to participate, including length of survey response time and

165 the number of potential years of involvement.

166 *Follow-up survey*

167     Beginning in September 2016, Xinjiang's government launched a comprehensive

168 physical examination project within the region. All residents in Xinjiang can

169 participate in this free annual physical examination, with the cost covered by the local

170 government. The project will continue for decades. This free annual health

171 examination for all residents in Xinjiang included but were not limited to physical

172 examination, health-risk survey, biochemical blood examination, ultrasound, and X-

173 ray. During the past four years, Xinjiang has set up more than 2,800 physical

174 examination centres and invested a total of 3.90 billion yuan in the Xinjiang Universal

175 Health Examination. These Universal Health Examination are routinely attended by

176 more than 80% of adults residing in the region. More than 50 million people have

177 taken part in universal health examinations up to September 2018. Results of physical

178 examination have formed a complete longitudinal database. After the cohort baseline

179 survey, we can link with the database through individual identification (ID) numbers

180 to follow our participants. We have matched the baseline survey data of XMC

181 subjects with the personnel database of the national physical examination within

182 Xinjiang in 2019. More than 70% of subjects in XMC participated in the national

183 physical examination within Xinjiang in 2019.

184     Participants' follow-up will conduct by recording unique ID numbers associated

185 with everyone across several common data systems. These data systems include

186 Medical record information system, Medical insurance management system, Maternal

187 and Child health information system, Public security household registration

188 management systems, Civil affairs management system, Chronic disease management

189 information system, National central cancer registry of china, and Death registration

190 information management system. Data from these different sources can be used

191 concordantly for mutual review and for supplementing survey data.

192     In the final year of cohort building, 10% of subjects were randomly selected for

193 telephone or face-to-face follow-up. Study participants given priority for this follow-

194 up included those that had not had any measurable outcome during the study period.

195 There are two main purposes for this strategy. Firstly, we can reduce the loss of

196 follow-up as much as possible. Secondly, we can check whether the medical database

197 can cover all participants by comparing the results of telephone or face-to-face

198 follow-up with the medical database matching.

199     As planned, we will revisit the study subjects on site every 2-3 years. Again, we

200 will conduct face-to-face questionnaires and collect biological specimens such as

201 blood and urine from the study subjects. We have completed a partial revisit of the

202 study population in 2021.

203 *Data and blood samples collection*

204     The data collection methods applied for the Xinjiang multi-ethnic cohort study

205 included a questionnaire survey, physical examination, and collection of biological

206 samples; measures which are essential for any longitudinal study. The questionnaire

207 survey was conducted by medical students who had received consent training, and

208 physical examinations were conducted by trained nurses or doctors.

209     The baseline questionnaire we used was mainly referred to as the baseline

210 questionnaire of CKB[8] and has been slightly modified according to the opinion of

211 experts from the medical colleges in Northwest China. The questionnaire collected

212 information on sociodemographics, tea and coffee consumption, alcohol intake,

213 tobacco use, dietary habits, passive smoking and indoor air pollution, personal and

214 family medical history, physical activity, mental health, and reproductive history

215 (female study participants only; Table 1). But our collaborators conducted an

216 additional dietary survey of the study participants in the Ili region, the data collected

217 on dietary habits included specific information as to the types of food consumed, as

218 well as the frequency and intake of specific food types. For example, information on

219 the frequency and intake amount of fruit consumed, as well as specifics pertaining to

220 fruit type, i.e., number of apples, bananas, and oranges consumed, was collected.

221

222    Table 1. Questionnaire data collected in the Xinjiang multi-ethnic cohort study

| Questionnaire survey | Description |
| --- | --- |
| Demographic data | Name, Gender, ID card, Medical insurance, Date of birth, Education level, Marital status, Occupation, Financial income |
| Tea and coffee consumption | Frequency of tea drinking in the past year, Previous tea drinking habits, Types of tea drinking, Frequency of coffee consumption |
| Alcohol intake | Frequency and amount of drinking in the past year, previous drinking habits (drinking refers to drinking in the past year under normal circumstances, not holidays or special periods, such as during marriage) |
| Tobacco use | Frequency and amount of smoking in the past year, previous smoking habits (current smoker defined as daily or almost daily smoking; former smoker defined as stopping smoking for at least 6 months; non-smoker defined as that the cumulative amount of smoking in lifetime does not exceed 100) |
| Dietary habits | Intake of 30 common foods, including staple foods, animal and plant foods, soy products, dairy products, vegetables, and fruits; Types of daily consumption oil; Use of nutrient supplements; Spicy food and vinegar consumption; Household refrigerator use time |
| Passive smoking and indoor air pollution | Whether had history of live with current smoker; Frequency of weekly exposure to passive smoking and cumulative exposure time; Cooking, heating, household fuel use, and house decoration situation; Occupational exposure history of air pollutants |
| Personal and family medical history | Self-health evaluation; chest and respiratory symptoms; history of 18 common diseases including diabetes, acute myocardial infarction, hypertension, asthma, and malignant tumors; history of blood transfusion and constipation; history of 5 common chronic diseases among family members |
| Physical activity | Work-related physical activity in the past year; the way of commuting and time spent on commuting; amateur physical exercise; housework activities; frequency and length of physical activity (i.e., sweating, heartbeat, etc.); weight loss in the past year |
| Mental health | Satisfaction with current living conditions; 10 major events that may have a serious impact on life in the past two years; sleep conditions (including insomnia, daily sleep time, whether to snoring while sleeping, etc.); depression, anxiety disorders, and unexplained phobias; quality of life |
| Reproductive history (only for females) | Menstrual history of female (age at menopause and menarche), pregnancy, fertility, breastfeeding, contraceptive use, and surgery |

223    Data collected during the physical examination included height (using a medical

224    height gauge with an accuracy of 0.1cm), weight (using a medical electronic scale

225    with an accuracy of 0.1kg), waist circumference (using a soft measuring tape with an

226  accuracy of 0.1 cm; the lower edge of the measuring tape was placed at the highest

227  points on both sides of the hip bone, and horizontal measurement of the waist

228  circumference taken), heart rate and blood pressure (measured using a medical arm

229  electronic blood pressure monitor with an accuracy of 2 mmHg; two measurements

230  per study participant were taken, following a rest period of five to ten minutes), body

231  fat composition (measured using a body composition analyzer, TANITA DC-

232  430MA ). Participants were asked to remove jackets, shoes, and hats while all

233  physical measurements were being taken.

234      A vacuum blood collection device with intravenous anticoagulant was used to

235  collect a 20 ml blood sample for each participant. A 4 ml blood sample was used for

236  biochemical examination and routine blood examination. These examinations were

237  completed at the nearest township health service center to the survey location. Whole

238  blood samples (3 ml) were transferred to three cryopreservation tubes immediately

239  after blood samples were collected. Blood samples used to separate plasma and white

240  blood cells were centrifuged within two hours of blood sample collection (4℃ at

241  3000 rpm for 10 minutes). Samples were stored at -196℃ in liquid nitrogen

242  containers before and after bi-monthly transportation to Urumqi. RNA protection

243  solution was also added to some samples for subsequent RNA detection. All the blood

244  and urine samples are placed in a special biological sample bank, which is managed

245  by a specially assigned person using an electronic management system.

246      For a large study such as this to be practicable and economically feasible in a

247  resource-poor setting, study procedures needed to be simple and streamlined. In this

248  study, the full assessment carried out at recruitment, including obtaining informed

249  consent, administering the questionnaire survey, carrying out the physical

250  examination, and obtaining blood samples, took an average of 60 minutes to

251  complete. Investigators from all three survey locations were trained by the same

252  junior investigator and postgraduate of the Xinjiang Medical University. The survey

253  equipment was also consistent across all three survey locations and equipment

254  calibration was carried out every day.

255  *Statistical analyses*

256  For the baseline profile, descriptive statistics were calculated for baseline data

257  regarding demographic data, tea and coffee consumption, alcohol intake, tobacco use,

258  dietary habits, passive smoking and indoor air pollution, personal and family medical

259  history, physical activity, mental health, reproductive history (only for female). All

260  analyses were conducted using SAS version 9.4.

261  **Results**

262  When a participant came to the assessment center and met our eligibility criteria,

263  they were first given a serial number, and a total of 31,778 participants were given a

264  number, but some of them dropped out during the questionnaire or refused to collect a

265  biological specimen. Finally, a total of 30,949 participants were recruited as part of

266  the Xinjiang multi-ethnic cohort study from Urumqi, Hotan, and Ili in Xinjiang,

267  China. The number of participants recruited from Urumqi, Hotan, and Ili was 10017,

268  12921, and 8011, respectively (Table 2). The average age of study participants was

269  56.21 years for men, and 54.75 years for women. In Urumqi, more than half of the

270    subjects were aged over 65 years old, significantly higher than the other two regions.

271    According to the Census data of Xinjiang[9], Urumqi has the highest proportion of

272    people aged 65 and over compared to the other two regions. Also, our baseline survey

273    was conducted on a weekday, but in urban areas, younger people are more likely to be

274    required to work on a weekday. The proportion of Han Chinese in the Urumqi study

275    population was close to 90%, while the proportion of Uyghur in the Hotan study

276    population was over 99%. The proportion of ethnic groups represented in the study

277    population recruited in the Ili region, conversely, was relatively balanced, indicating

278    that ethnic composition varies significantly from region to region. The average level

279    of highest educational attainment and average income among the Urumqi population

280    was significantly higher than that of the Hotan and Ili populations. The proportion of

281    people who had been educated to primary or pre-primary level in Hotan and Yili were

282    87.65% and 71.14%, respectively.

283        Table 2. Demographic characteristics of the study participants by site

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Men** | **Women** | **Total** | **Men** | **Women** | **Total** | **Men** | **Women** | **Total** |
| Age(years) | | | | | | | | | |
| <45 | 231(5.87) | 330(5.43) | 561(5.60) | 1040(21.17) | 2337(29.18) | 3377(26.14) | 1120(29.60) | 1250(29.57) | 2370(29.58) |
| 45-54 | 570(14.47) | 1021(16.80) | 1591(15.88) | 1408(28.66) | 2871(35.85) | 4279(33.12) | 1248(32.98) | 1537(36.36) | 2785(34.76) |
| 55-64 | 824(20.92) | 1358(22.34) | 2182(21.78) | 1364(27.76) | 1848(23.08) | 3212(24.86) | 870(22.99) | 994(23.52) | 1864(23.27) |
| 65- | 2313(58.74) | 3370(55.44) | 5683(56.73) | 1101(22.41) | 952(11.89) | 2053(15.89) | 546(14.43) | 446(10.55) | 992(12.38) |
| Ethnicity | | | | | | | | | |
| Han | 3457(87.79) | 5432(89.36) | 8889(88.74) | 6(0.12) | 6(0.07) | 12(0.09) | 572(15.12) | 522(12.35) | 1094(13.66) |
| Hui | 310(7.87) | 392(6.45) | 702(7.01) | 4(0.08) | 15(0.19) | 19(0.15) | 1336(35.31) | 1333(31.54) | 2669(33.32) |
| Uyghur | 105(2.67) | 168(2.76) | 273(2.73) | 4897(99.67) | 7969(99.51) | 12866(99.57) | 908(24.00) | 1361(32.20) | 2269(28.32) |
| Kazakh | 41(1.04) | 53(0.87) | 94(0.94) | 6(0.12) | 13(0.16) | 19(0.15) | 883(23.34) | 951(22.50) | 1834(22.89) |
| Other | 25(0.63) | 34(0.56) | 59(0.59) | NA | 5(0.06) | 5(0.04) | 85(2.25) | 60(1.42) | 145(1.81) |
| Education | | | | | | | | | |
| Primary/less | 1064(27.02) | 2534(41.68) | 3598(35.92) | 4093(83.31) | 7232(90.31) | 11325(87.65) | 2582(68.23) | 3117(73.74) | 5699(71.14) |
| Secondary | 2320(58.91) | 3020(49.68) | 5340(53.31) | 774(15.75) | 734(9.17) | 1508(11.67) | 1166(30.81) | 1071(25.34) | 2237(27.92) |
| Degree or above | 553(14.04) | 517(8.50) | 1070(10.68) | 32(0.65) | 18(0.22) | 50(0.39) | 28(0.74) | 26(0.62) | 54(0.67) |
| Refused/missing | 1(0.03) | 8(0.13) | 9(0.09) | 14(0.28) | 24(0.30) | 38(0.29) | 8(0.21) | 13(0.31) | 21(0.26) |

Marriage

15

| | Urumqi (n=10017) | | | Hotan (n=12921) | | | Ili (n=8011) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Men** | **Women** | **Total** | **Men** | **Women** | **Total** | **Men** | **Women** | **Total** |
| Married | 3696(93.85) | 5045(82.99) | 8741(87.26) | 4319(87.91) | 6165(76.99) | 10484(81.14) | 3500(92.49) | 3586(84.84) | 7086(88.45) |
| Widowed/separated | 197(5.00) | 1006(16.55) | 1203(12.01) | 538(10.95) | 1773(22.14) | 2311(17.89) | 227(6.00) | 614(14.53) | 841(10.50) |
| Never married | 42(1.07) | 19(0.31) | 61(0.61) | 19(0.39) | 6(0.07) | 25(0.19) | 43(1.14) | 6(0.14) | 49(0.61) |
| Refused/missing | 3(0.08) | 9(0.15) | 12(0.12) | 37(0.75) | 64(0.80) | 101(0.78) | 14(0.37) | 21(0.50) | 35(0.44) |
| Income (RMB)* | | | | | | | | | |
| under 10000 | 632(16.05) | 912(15.00) | 1544(15.41) | 2689(54.73) | 5139(64.17) | 7828(60.58) | 414(10.94) | 616(14.57) | 1030(12.86) |
| 10000-19999 | 412(10.46) | 585(9.62) | 997(9.95) | 1196(24.34) | 1649(20.59) | 2845(22.02) | 1311(34.65) | 1712(40.50) | 3023(37.74) |
| 20000-34999 | 411(10.44) | 737(12.12) | 1148(11.46) | 453(9.22) | 501(6.26) | 954(7.38) | 1157(30.58) | 1093(25.86) | 2250(28.09) |
| 35000-49999 | 504(12.80) | 885(14.56) | 1389(13.87) | 148(3.01) | 198(2.47) | 346(2.68) | 393(10.39) | 339(8.02) | 732(9.14) |
| 50000 or above | 1978(50.23) | 2952(48.56) | 4930(49.22) | 416(8.47) | 500(6.24) | 916(7.09) | 503(13.29) | 459(10.86) | 962(12.01) |
| Refused/missing | 1(0.03) | 8(0.13) | 9(0.09) | 11(0.22) | 21(0.26) | 32(0.25) | 6(0.16) | 8(0.19) | 14(0.17) |
| Anthropometric index | | | | | | | | | |
| Height | 169.03±6.26 | 157.98±19.33 | 162.33±16.47 | 163.32±7.29 | 153.60±6.81 | 157.30±8.44 | 167.13±6.98 | 155.98±7.05 | 161.24±8.95 |
| Weight | 73.82±10.69 | 63.10±11.46 | 67.31±12.33 | 66.71±12.38 | 59.63±12.04 | 62.32±12.64 | 73.24±12.06 | 65.20±12.33 | 68.52±12.69 |
| Waist circumference | 92.40±17.00 | 87.09±15.47 | 89.18±16.30 | 91.88±11.50 | 91.44±12.29 | 91.61±12.00 | 90.56±11.02 | 88.49±11.80 | 89.46±11.48 |
| Body mass index | 25.81±3.31 | 25.35±4.42 | 25.53±4.02 | 24.96±4.04 | 25.24±4.67 | 25.14±4.44 | 25.82±3.75 | 26.76±4.51 | 26.31±4.20 |

284    N(%) for categorical variable; Mean ± SD for continuous variables; *: 1 RMB = 0.14524 USD and 1 RMB = 0.1277 EUR in December 2018.

16

285    Approximately 50% of participants surveyed in Urumqi had an average annual

286    income of more than 50,000, whereas about 60% of subjects in Hotan had an average

287    annual income of less than 10,000. Study participants in Urumqi were found to have

288    the highest average height, while study participants in Ili were found to have the

289    highest average weight, with study participants in Hotan having the highest average

290    waist circumference. The average body mass index of the subjects in the three regions

291    exceeded the diagnostic threshold (BMI $\geq$ 25kg/m$^2$)[10] for overweight in China,

292    with study participants in the Ili region having the highest average body mass index.

293      The five most common chronic diseases among participants across all three

294    survey sites were dyslipidemia, hypertension, cholecystitis, diabetes, ischaemic heart

295    disease with a prevalence of 34.55%, 32.66%, 14.34%, 10.07%, and 8.74%,

296    respectively (Table 3).

297    Table 3. Self-reported disease prevalence among subjects in different sites (%)

| Diseases | Urumqi | | Hotan | | Ili | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Crude | Adjusted | Crude | Adjusted | Crude | Adjusted | Crude | Adjusted |
| Cardiovascular diseases | | | | | | | | |
| Hypertension* | 47.19 | 29.29 | 37.00 | 32.64 | 38.61 | 36.09 | 40.74 | 32.66 |
| Hypertension | 32.64 | 17.87 | 27.35 | 23.76 | 16.71 | 15.23 | 26.31 | 19.80 |
| IHD | 8.18 | 3.28 | 13.54 | 12.15 | 8.44 | 7.72 | 10.48 | 8.74 |
| Other heart disease | 0.57 | 0.32 | 2.63 | 2.34 | 0.69 | 0.62 | 1.46 | 1.27 |
| Stroke | 3.43 | 1.62 | 4.91 | 4.31 | 1.61 | 1.47 | 3.58 | 2.82 |
| Metabolic disease and kidney diseases | | | | | | | | |
| Diabetes* | 23.93 | 19.14 | 7.26 | 6.74 | 9.80 | 9.19 | 13.11 | 10.07 |
| Diabetes | 14.73 | 8.81 | 4.25 | 3.87 | 4.69 | 4.36 | 7.76 | 5.54 |
| Dyslipidemia* | 35.00 | 35.08 | 33.53 | 34.05 | 34.83 | 35.08 | 34.32 | 34.55 |
| CKD | 0.68 | 0.56 | 6.20 | 6.28 | 0.99 | 0.93 | 3.06 | 3.17 |
| Anemia* | 3.31 | 5.30 | 5.89 | 5.55 | 5.30 | 5.26 | 4.98 | 5.44 |
| Chronic respiratory diseases | | | | | | | | |

17

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Chronic bronchitis | 4.48 | 2.16 | 15.89 | 14.17 | 3.31 | 3.12 | 8.94 | 7.87 |
| Emphysema | 0.29 | 0.15 | 3.00 | 2.57 | 0.51 | 0.49 | 1.48 | 1.25 |
| COPD | 0.40 | 0.17 | 1.22 | 1.14 | 0.20 | 0.18 | 0.69 | 0.62 |
| Asthma | 1.11 | 0.65 | 2.86 | 2.45 | 0.22 | 0.20 | 1.61 | 1.36 |
| Tuberculosis | 0.69 | 0.57 | 2.84 | 2.55 | 0.54 | 0.50 | 1.55 | 1.37 |
| Digestive diseases | | | | | | | | |
| Chronic hepatitis | 0.49 | 0.53 | 4.85 | 4.91 | 1.34 | 1.33 | 2.53 | 2.82 |
| Cholecystitis | 6.50 | 3.12 | 23.28 | 21.28 | 13.66 | 12.49 | 15.36 | 14.34 |
| Peptic ulcer | 0.91 | 0.74 | 6.61 | 6.33 | 2.11 | 2.01 | 3.60 | 3.61 |
| Skeletal disorders | | | | | | | | |
| Osteoporosis | 2.15 | 0.83 | 7.48 | 6.50 | 3.74 | 3.52 | 4.79 | 4.21 |
| Fracture | 1.93 | 1.24 | 5.82 | 6.10 | 5.75 | 5.87 | 4.54 | 4.96 |
| Cancer | 0.63 | 0.32 | 0.67 | 0.54 | 0.46 | 0.43 | 0.60 | 0.48 |

298    *Self-reported combined with biological detection

299    Adjusted: Age and gender adjusted based on 2010 China census data

300    IHD: Ischaemic heart disease, Other heart disease: included Rheumatic heart disease and Cor pulmonale,

301    CKD: Chronic kidney disease

302    Disease prevalence varies substantially across the three survey locations, which

303    may be attributed to differences in nationality, diet, education, socioeconomics, and

304    living habits between the three areas. After standardizing for age and sex based on

305    2010 China census data, the prevalence of hypertension in Ili was 36.09%, higher than

306    that in Urumqi (29.29%) and Hotan (32.64%). According to a survey carried out

307    between 2012 to 2015 the average prevalence of hypertension in China was 23%[11].

308    The prevalence of hypertension in the XMC study sites, as determined by this study,

309    was much higher than the national average. Study participants in Urumqi had a higher

310    prevalence of diabetes (19.14%) compared with those in Ili and Hotan. The results of

311    this study also found that the prevalence of diabetes in Ili and Hotan was slightly

312    lower than the national average (10.9%).[12] Urumqi has a high prevalence of

313    diabetes and hypertension, which may be associated with participants residing in

18

314 urban communities and having a higher average income compared with rural dwelling

315 populations in Hotan and Ili. As a result, Urumqi's people have better health resources

316 so people with diabetes are more likely to be diagnosed correctly. In addition,

317 Urumqi's people have a lower proportion of the physical labor workforce than the

318 other two regions, so that they may be sedentary in work time, which may lead to a

319 higher prevalence of diabetes in the Urumqi population. The prevalence of chronic

320 kidney disease among participants in Hotan (6.28%) was approximately ten times

321 higher than that of Urumqi (0.56%) and six times that of Ili (0.93%). The prevalence

322 of cholecystitis in Hotan (21.28%) was approximately seven times that of the

323 prevalence in Urumqi (3.12%). Results of this study also found that the prevalence of

324 COPD in Xinjiang (0.62%) was much lower than the national level (3.84%)[13].

325 Regarding Cholecystitis, the data on the national average was limited. A previous

326 epidemiological survey showed that the incidence of gallstones in Xinjiang was

327 11.83%.[14] The prevalence of Cholecystitis in Ili (12.49%) and Hotan (21.28%) was

328 much higher than previously recorded levels. This may be associated with Hotan's

329 local dietary habits. The prevalence of ischemic heart disease, stroke, chronic

330 hepatitis, and CKD in Hotan was also found to be higher than that of the other two

331 sites.

332     About baseline dietary intake (see Figure 3), more than 60% of participants in all

333 three survey locations reported an average consumption of fruit and vegetable three or

334 more times per week. In Hotan and Ili, the staple food was wheaten food, whereas, in

335 Urumqi, rice and wheaten food was the food staples. Dietary habits in Hotan, which

19

336    has a predominantly Uyghur population, consisted predominantly of mutton, with

337    more than 70% consuming mutton three or more times per week. The proportion of

338    the population who consumed pork, mutton, or beef three or more times per week in

339    Urumqi by contrast, where the population is predominantly Han, was 28.26%, 20.24,

340    and 24.28%, respectively. Consumption of white meat, such as fish and poultry, was

341    lower in the three survey locations. Fresh meat consumed by Xinjiang residents

342    consisted mainly of red meat such as beef and mutton. Some studies have suggested

343    that too high an intake of red meat might increase the risk of metabolic syndrome

344    [15,16]. The study also showed that consumption of fresh milk was associated with a

345    lower risk of Metabolic Syndrome[16]. In this study, the proportion of people

346    drinking milk three or more times per week was lower in Ili and particularly in Hotan

347    (<10% of participants surveyed). More than 60% of people in Ili reported drinking

348    milk tea ≥3 times a week, however, with few of them drinking fresh milk directly.

349        A diet high in salt and fat is associated with an increased risk of high blood

350    pressure and ischemic heart disease[17]. The consumption of coarse grains, eggs, milk

351    and yogurt three or more times per week among people surveyed in Urumqi was

352    higher than among people surveyed in Ili and Hotan. In addition to this, consumption

353    of fowl, fish or seafood, and beans three or more times per week was low in all three

354    survey locations (<10% of participants surveyed). The results of the survey on dietary

355    habits reported here are generally consistent with a recent survey[18], which found

356    that consumption of cereals, meat, and oil in Xinjiang exceeded maximum

357    recommended allowances, while consumption of eggs, seafood products, dairy

20

358 products, nuts, and legumes were below the minimum recommended allowances.

359 Notably, this study reported higher consumption of fruit and vegetable in all three

360 regions than previous studies on fruit and vegetable consumption in Xinjiang[16,19].

361 **Discussion**

362     This is the largest prospective study that has been undertaken to date in Xinjiang,

363 China. The main strength of the Xinjiang multi-ethnic cohort study is the inclusion of

364 over 20,000 survey participants from diverse ethnic backgrounds such as Uyghur,

365 Kazakh, and Hui, which allows prevalence of chronic conditions, as well as dietary

366 habits, various health determinants, and socio-demographics between distinct

367 populations to be compared. It is already well established that the dietary habits of

368 Uyghur, Kazakh, and Hui differ from those of Han Chinese. We have investigated the

369 dietary habits of all the study participants as part of this cohort study, in conjunction

370 with prevalence of non-communicable diseases to compare health determinants and

371 prevalence of chronic conditions between these three survey locations.

372     Another important strength of this study was the survey questionnaire, which

373 collected details on major risk factors such as alcohol intake, tobacco use (we

374 investigated current as well as past usage quantity, frequency, type), mental health,

375 and quality of life. The three survey locations selected for this study were broadly

376 representative of the different populations residing in Xinjiang, including animal

377 herders, and urban and rural residents. The results of the survey conducted here can be

378 used to compare differences in prevalence and type of non-communicable chronic

379 diseases among these three populations.

21

380    There were also several limitations with the Xinjiang multi-ethnic cohort study.

381  Firstly, we did not use random sampling to select study participants, but rather a

382  convenience sampling strategy whereby eligible study participants from

383  predetermined survey locations were recruited until an adequate sample size had been

384  met. We excluded those who had lived in the local area for less than one year, which

385  may reduce how representative the study population is for Urumqi regions. However,

386  for the Ili and Hotan regions, the population is more stable and may not have an

387  impact on the representation of the population. As this study is the first large-scale

388  cohort study to have been carried out in Xinjiang, and data collected as part of

389  previous cross-sectional health surveys of Xinjiang residents have not yet been

390  published, we cannot compare differences in the distribution of diseases and related

391  factors between our cohort study population and the wider Xinjiang population.

392  Another limitation is that diseases and exposures were self-reported. Although a

393  detailed and comprehensive questionnaire was carried out, some results are subject to

394  reporting and recall bias.

395    We will take advantage of the existing monitoring systems or database of

396  Xinjiang, China to obtain information on health outcomes (such as morbidity and

397  death events) for all participants of the cohort study. Repeated cross-sectional surveys

398  of study participants will be conducted on a bi-annual basis with surveys focusing on

399  research outcome being the primary concern. Data will be obtained by administering

400  questionnaires used for baseline research, as well as additional information on health

401  determinants. Although the study population is relatively large, the availability of

22

402  repeated measures, extensive biobank blood, and urine samples, and health

403  information systems linked by an ID number will allow this study to assess within-

404  and between-person variability over time in major risk factors among the population

405  of people living in Xinjiang, China.

406  **Contributorship statement:** Conceptualization: Hong Yan, Jianghong Dai;

407  Methodology: Jianghong Dai, Hong Yan, Qi Sun, Gaofeng Sun; Software: Tao Luo;

408  Validation: Tian Tian; Formal analysis: Tao Luo, Tian Tian; Investigation: Tao Luo,

409  Zewen Zhang, Liu, Lirong, Qi Sun, Gaofeng Sun; Resources: Jianghong Dai, Hong

410  Yan; Data curation: Tao Luo, Zewen Zhang, Liu, Lirong; Writing-original draft

411  preparation: Tao Luo, Tian Tian, Zewen Zhang, Liu, Lirong; Writing-review and

412  editing: Jianghong Dai; Visualization: Tao Luo; Supervision: Jianghong Dai; Project

413  administration: Jianghong Dai, Hong Yan; Funding acquisition: Jianghong Dai, Hong

414  Yan.

415  **Competing interests:** No competing interest.

420  **Data sharing statement:** Baseline data from the Xinjiang multi-ethnic cohort study

421  are put through a process of editing. The study data are not freely available, but

422  specific proposals for future collaborations are welcome. Address to the research

423  leader of Xinjiang multi-ethnic cohort study.

23

426    **References**

427    1    Chen Z, Chen J, Collins R, *et al.* China Kadoorie Biobank of 0.5 million people: survey methods,
428         baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;**40**:1652–66.
429         doi:10.1093/ije/dyr120

430    2    Zhang T, Zhang C, Zhang Y, *et al.* Metabolic syndrome and its components as predictors of
431         nonalcoholic fatty liver disease in a northern urban Han Chinese population: a prospective
432         cohort study. *Atherosclerosis* 2015;**240**:144–8. doi:10.1016/j.atherosclerosis.2015.02.049

433    3    Wang X, Lu M, Qian J, *et al.* Rationales, design and recruitment of the Taizhou Longitudinal
434         Study. *BMC Public Health* 2009;**9**:223. doi:10.1186/1471-2458-9-223

435    4    Song K, Du H, Zhang Q, *et al.* Serum immunoglobulin M concentration is positively related to
436         metabolic syndrome in an adult population: Tianjin Chronic Low-Grade Systemic
437         Inflammation and Health (TCLSIH) Cohort Study. *PLoS ONE* 2014;**9**:e88701.
438         doi:10.1371/journal.pone.0088701

439    5    Li L, Lyu J. Large prospective cohort studies: a review and update. *Zhonghua Liu Xing Bing
440         Xue Za Zhi* 2015;**36**:1187–9.

441    6    He J, Guo S, Liu J, *et al.* Ethnic differences in prevalence of general obesity and abdominal
442         obesity among low-income rural Kazakh and Uyghur adults in far western China and
443         implications in preventive public health. *PLoS ONE* 2014;**9**:e106723.
444         doi:10.1371/journal.pone.0106723

445    7    ZHE W, PUERHATI W, LIU L, *et al.* Analysis of the death causes of residents in 15
446         surveillance sites in Xinjiang in 2015. *Modern Preventive Medicine* 2017;**44**:4186–90.

447    8    Li L, Lv J, Guo Y, *et al.* The China Kadoorie Biobank: related methodology and baseline
448         characteristics of the participants. *Zhonghua Liu Xing Bing Xue Za Zhi* 2012;**33**:249–55.

449    9    Office of the Leading Group of the Xinjiang Uygur Autonomous Region for the Seventh
450         National Population Census. Communiqué of the Seventh National Population Census of the
451         Xinjiang Uygur Autonomous Region. Statistic Bureau of Xinjiang Uygur Autonomous Region
452         2021.

453    10   Reynolds K, Gu D, Whelton PK, *et al.* Prevalence and risk factors of overweight and obesity
454         in China. *Obesity (Silver Spring)* 2007;**15**:10–8. doi:10.1038/oby.2007.527

455    11   Wang Z, Chen Z, Zhang L, *et al.* Status of Hypertension in China: Results From the China
456         Hypertension    Survey,    2012-2015.    *Circulation*    2018;**137**:2344–56.
457         doi:10.1161/CIRCULATIONAHA.117.032380

24

458    12  Wang L, Gao P, Zhang M, *et al.* Prevalence and Ethnic Pattern of Diabetes and Prediabetes in
459        China in 2013. *JAMA* 2017;**317**:2515–23. doi:10.1001/jama.2017.7596

460    13  Chan KY, Li X, Chen W, *et al.* Prevalence of chronic obstructive pulmonary disease (COPD)
461        in China in 1990 and 2010. *J Glob Health* 2017;**7**:020704. doi:10.7189/jogh.07.020704

462    14  Zhu L, Aili A, Zhang C, *et al.* Prevalence of and risk factors for gallstones in Uighur and Han
463        Chinese. *World J Gastroenterol* 2014;**20**:14942–9. doi:10.3748/wjg.v20.i40.14942

464    15  de Oliveira Otto MC, Alonso A, Lee D-H, *et al.* Dietary intakes of zinc and heme iron from red
465        meat, but not from other sources, are associated with greater risk of metabolic syndrome and
466        cardiovascular disease. *J Nutr* 2012;**142**:526–33. doi:10.3945/jn.111.149781

467    16  Guo H, Gao X, Ma R, *et al.* Prevalence of Metabolic Syndrome and its Associated Factors
468        among Multi-ethnic Adults in Rural Areas in Xinjiang, China. *Sci Rep* 2017;**7**:17643.
469        doi:10.1038/s41598-017-17870-5

470    17  Zhai FY, Du SF, Wang ZH, *et al.* Dynamics of the Chinese diet and the role of urbanicity,
471        1991-2011. *Obes Rev* 2014;**15 Suppl 1**:16–26. doi:10.1111/obr.12124

472    18  Yin J, Yang D, Zhang X, *et al.* Diet shift: Considering environment, health and food culture.
473        *Sci Total Environ* 2020;**719**:137484. doi:10.1016/j.scitotenv.2020.137484

474    19  Zhai F, He Y, Wang Z, *et al.* Status and characteristic of dietary intake of 12 minority
475        nationalities in China. *Wei Sheng Yan Jiu* 2007;**36**:539–41.

476    Figure legends
477    Table 1. Questionnaire data collected in the Xinjiang multi-ethnic cohort study
478    Table 2. Demographic characteristics of the study participants by site
479    Table 3. Self-reported disease prevalence among subjects in different sites
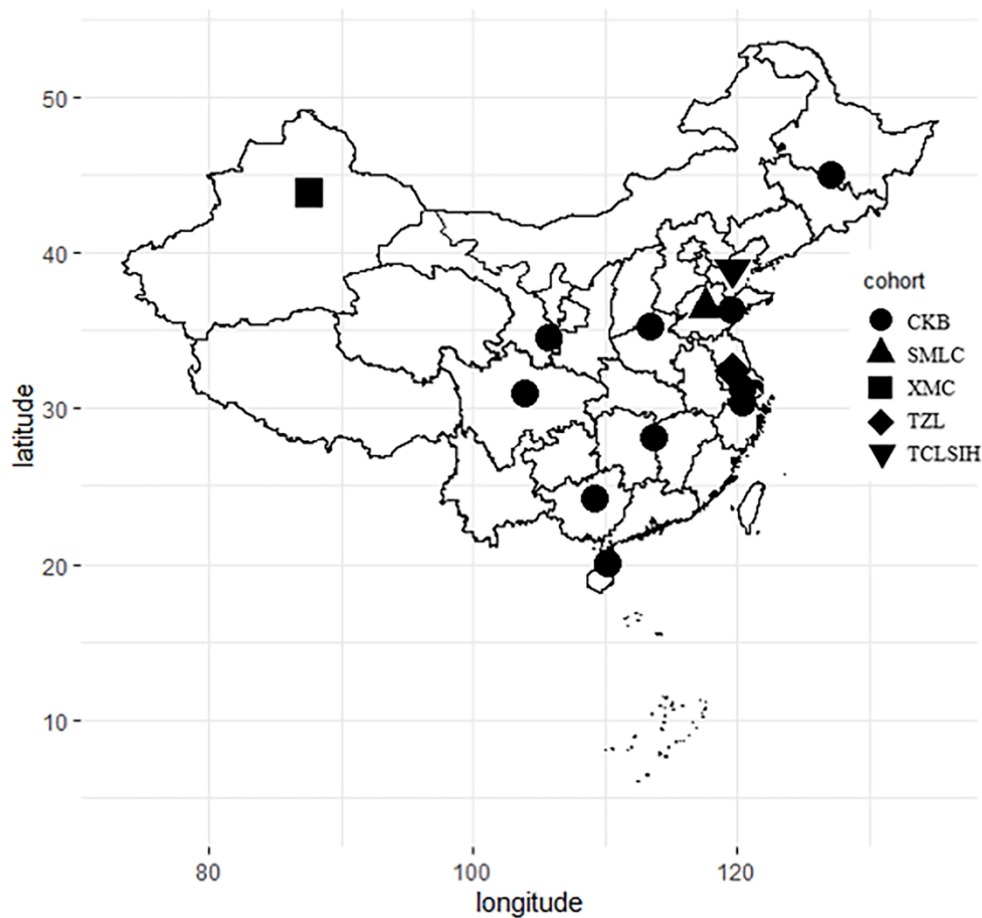480    Figure 1. Locations of on-going prospective cohorts in China
481    CKB: China Kadoorie Biobank; SMLC: Shandong Multi-center Longitudinal Cohort; XMC: Xinjiang Multi-ethnic Cohort; TZL: Taizhou
482    Longitudinal Study; TCLSIH: Tianjin Chronic Low-Grade Systemic Inflammation and Health Cohort Study
483    Figure 2. Map of Xinjiang, China, showing the locations of three surveyed sites (Red five-pointed star)
484    in the XMC
485    Figure 3. Percentage of common food intake in different sites in the XMC
486

25

Figure 1. Locations of on-going prospective cohorts in China.CKB: China Kadoorie Biobank; SMLC: Shandong Multi-center Longitudinal Cohort; XMC: Xinjiang Multi-ethnic Cohort; TZL: Taizhou Longitudinal Study; TCLSIH: Tianjin Chronic Low-Grade Systemic Inflammation and Health Cohort Study
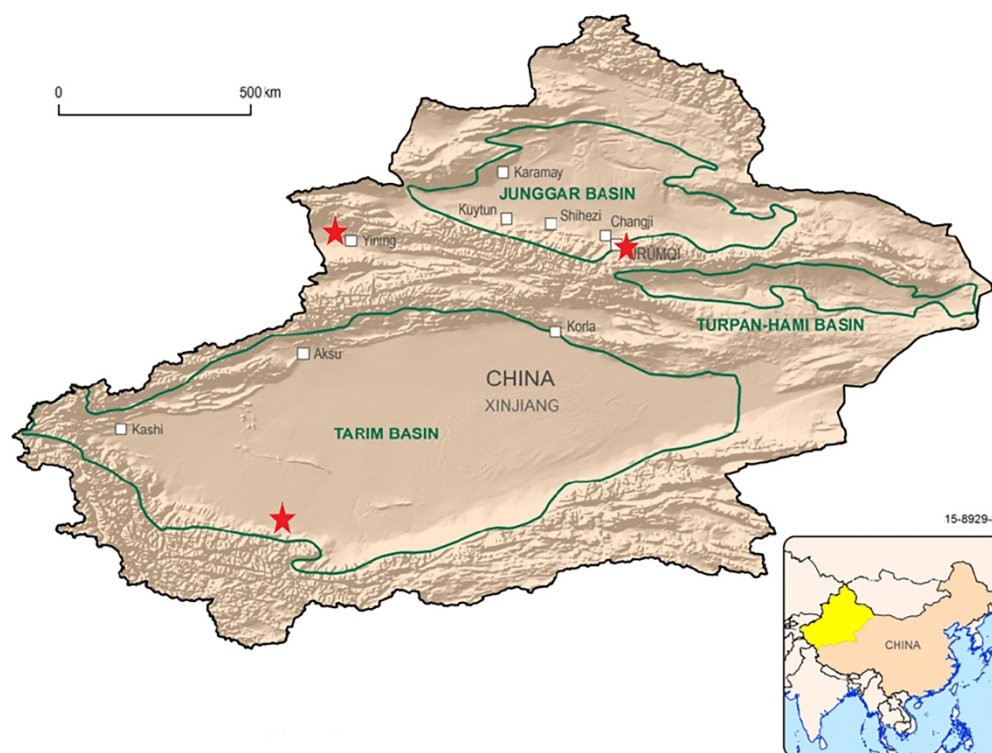
296x274mm (300 x 300 DPI)

Figure 2. Map of Xinjiang, China, showing the locations of three surveyed sites (Red five-pointed star) in the XMC
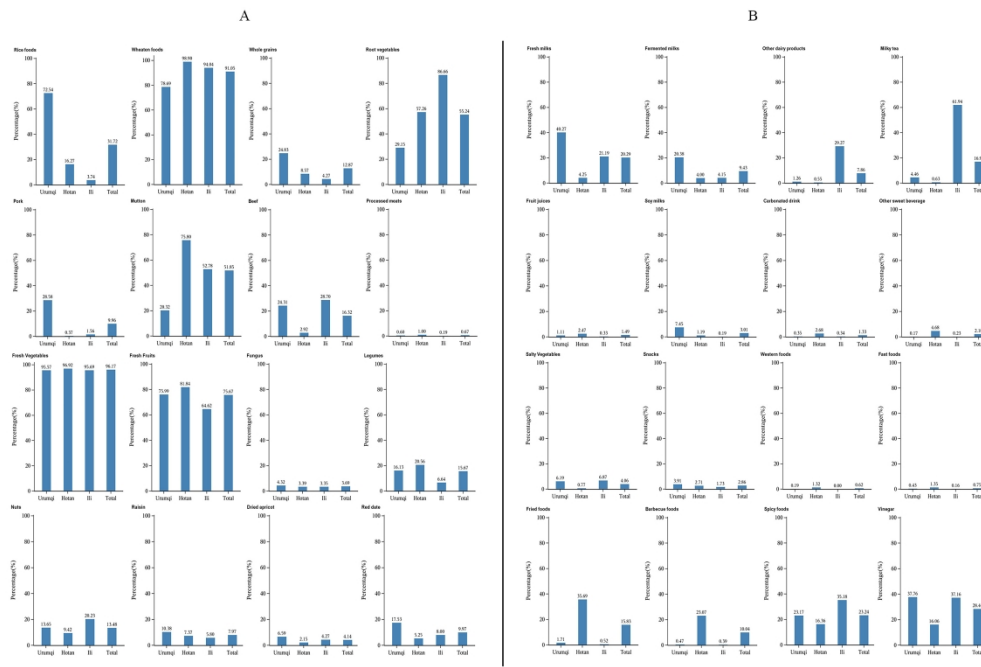
571x430mm (300 x 300 DPI)

Figure 3. Percentage of common food intake in different sites in the XMC

635x425mm (300 x 300 DPI)