

Design and Protocol of the Biobank for Metabolic Syndrome Consequences (BMSC): A Prospective Cohort Study in Northwest China

Xue Yang^{1,2,*}, Qian Li^{1-3,*}, Yayi He^{1,*}, Guozhi Yin⁴, Meng Li¹, Wenzhi Zhu², Xiayue Fan², Yi Gong², Yawen Wang⁵, Wei Qiang¹, Youfa Wang^{3,6}, Ko Willems van Dijk^{7,8}, Patrick CN Rensen^{2,8}, Hui Guo¹, Bingyin Shi¹, Yanan Wang¹⁻³

¹Department of Endocrinology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; ²Med-X Institute, Center for Immunological and Metabolic Diseases, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; ³International Obesity and Metabolic Disease Research Center, Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; ⁴Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; ⁵Biobank, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; ⁶Global Health Institute, School of Public Health, Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; ⁷Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; ⁸Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands

*These authors contributed equally to this work

Correspondence: Yanan Wang, Email yanan.wang@xjtu.edu.cn

Abstract: This manuscript describes the design and protocol of the Biobank for Metabolic Syndrome Consequences (BMSC), a prospective cohort study in Northwest China. Metabolic syndrome (MetS) is characterized by a group of interrelated disorders, including abdominal obesity, hyperglycemia, hypertension, and dyslipidemia. The presence of three or more of these conditions markedly increases the risk of multiple chronic diseases and mortality. The pathophysiology and natural course of MetS and its consequences are insufficiently understood. To improve our understanding, longitudinal research that combines biomarkers with longitudinal data measured over multiple time points is imperative. The BMSC, launched in August 2021 and still ongoing, is a prospective observational study of 2000 Chinese participants aged 18 to 75 years with MetS or relevant disorders living in the Northwest of China. At baseline survey, data on sociodemography, disease history, behavior and lifestyle, and mental health are collected by a structured questionnaire. The anthropometry is conducted by trained researchers. Fasting peripheral venous blood, urine, stool, and hair samples are collected according to standardized protocols. Extensive physical examinations are conducted in specific subgroups. Participants will be followed up every 3 months for at least 5 years for the incidence of MetS-related outcomes, such as cardiovascular disease, with clinical data and biological samples being collected at intervals similar to the baseline. These findings may contribute to improved prevention, early diagnosis, and personalized treatment of MetS-related conditions.

Keywords: metabolic syndrome, obesity, diabetes, biobank, study design

Background

Metabolic syndrome (MetS) is a cluster of chronic conditions, including abdominal obesity, hyperglycemia, hypertension, and dyslipidemia,¹ and has become a major global public health concern. The prevalence of MetS varies somewhat depending on the criteria used for the definition. A meta-analysis of global data from 28 million individuals reported a prevalence between 12.5% and 31.4% among adults worldwide.² In China, the prevalence of MetS grew from 8.8% in 1991–1995 to 29.3% in 2011–2015,³ and current estimates are up to 29.2% for men and even 35.4% for women,⁴ coinciding with the steep incline in the prevalence of obesity and diabetes. These high estimates necessitate a greater focus on MetS prevention.

MetS and its components are risk factors for cardiometabolic diseases (CMDs), such as type 2 diabetes (T2D), metabolic dysfunction-associated steatotic liver disease (MASLD), and cardiovascular diseases (CVDs)^{5–8} and contribute

to an increased mortality risk.^{8–10} Comorbidities such as chronic kidney disease (CKD),^{11,12} chronic obstructive pulmonary disease (COPD),^{13,14} depression,¹⁵ and cognitive impairment¹⁶ are also more prevalent in populations suffering from MetS. Given the enormous public health and economic burden of MetS, it is essential to understand the underlying pathophysiology and natural course of MetS and its consequences.

The etiology and mechanisms of MetS are heterogeneous and remain to be fully elucidated. Its development is influenced by modifiable risk factors, including obesity, sedentariness, and a high-fat diet.^{17,18} However, the pathways through which these factors contribute to MetS, and their interactions with genetic predisposition, are still not fully understood. Furthermore, evidence regarding the long-term dynamic changes in risk factors and their impact on the progression and consequences of MetS remains scarce, largely due to the limited availability of prospective cohorts with repeated measurements. Mechanistic studies have suggested that defective adipocyte differentiation and excessive visceral fat accumulation can drive the sustained release of proinflammatory cytokines, which may serve as primary triggers of MetS.^{19,20} Against this background, exploring novel biomarkers may provide further opportunities to better characterize the pathophysiology of MetS and its consequences; however, biomarkers suitable for early diagnosis of MetS-related diseases, such as metabolic dysfunction-associated steatohepatitis (MASH), are still lacking. Early identification of high-risk subgroups within populations with MetS is of great value for decreasing the risk of serious complications, underscoring the importance of developing such biomarkers.

Although many prospective cohorts have previously been established, they were largely community-based and focused broadly on the prevention of MetS and relevant metabolic disorders, with limited emphasis on in-depth mechanistic exploration. Several large-scale biobanks, such as UK Biobank in Europe, the China Kadoorie Biobank in China, FinnGen in Finland, and All of Us Research Program in the US, have provided valuable insights into the epidemiology, genetic architecture, and risk factors of chronic diseases, including metabolic disorders. Their success has demonstrated the power of biobank-based resources in advancing disease prevention and precision medicine. However, few of these cohorts have incorporated multiple repeated measurements during follow-up, particularly after the onset of MetS, which restricts the ability to capture dynamic changes over time. Moreover, none of these existing biobanks were specifically designed to investigate MetS as the primary focus, nor did they systematically integrate multi-omics profiling. In addition, dietary behavior is closely related to the occurrence and development of metabolic diseases. Northwest China is characterized by a high-carbohydrate diet and distinct lifestyle patterns, leading to a disease profiling that may differ from that of other regions. To date, no large-scale studies with repeated follow-up assessments and ample biological samples have been established for this specific population. The BMSC cohort was therefore established to fill this critical gap by combining longitudinal repeated measurements with multi-omics data in a uniquely underrepresented population.

To better understand the pathophysiology and natural course of MetS, it is imperative to combine studies with biomarkers and longitudinal data. Thus, we now aim to establish a comprehensive databank and biobank in Northwest China, i.e., the BMSC. This biobank is designed to 1) further understand the pathophysiology and consequences of MetS; 2) explore novel biomarkers to prevent and diagnose MetS and its consequences; and 3) supply useful information for the treatment of MetS and relevant disorders in a longitudinal observational study. By providing longitudinal multi-omics data in a well-characterized population, the BMSC may also inform personalized prevention strategies, enable early risk prediction, and guide targeted interventions for MetS-related conditions. This paper will describe the study design and the protocol of data collection in detail.

Study Population and Design

Study Population

Men and women aged 18 to 75 years are invited to participate in the BMSC study if they have been diagnosed with at least one of the following metabolic disorders: pre-diabetes or T2D, body mass index (BMI) ≥ 24.0 kg/m², hypertension, and/or dyslipidemia. The definitions of these metabolic disorders are based on established international guidelines and were applied uniformly to all participants. Specifically, pre-diabetes is defined as a fasting plasma glucose (PG) level of 5.6–6.9 mmol/L or a 2-hour PG level of 7.8–11.0 mmol/L after an oral glucose tolerance test (OGTT), while diabetes is defined as a fasting PG ≥ 7.0 mmol/L or 2-h PG ≥ 11.1 mmol/L or glycosylated hemoglobin (HbA1c) $\geq 6.5\%$.²¹

Hypertension is defined as systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or the current use of antihypertensive medication, in line with the Chinese Hypertension Guidelines (2020, in Chinese). Dyslipidemia is defined as meeting any of the following criteria: total cholesterol ≥ 5.2 mmol/L, low-density lipoprotein cholesterol (LDL-C) ≥ 3.4 mmol/L, triglycerides ≥ 1.7 mmol/L, or high-density lipoprotein cholesterol (HDL-C) < 1.0 mmol/L.²² Those who have lived locally for a long time (> 1 year) and are willing to accept long-term follow-up are eligible for the BMSC study. Individuals will be excluded if they: 1) are judged to have a life expectancy of less than 5 years; 2) are addicted to drugs or have a history of drug abuse; 3) have viral hepatitis, sexually transmitted diseases such as AIDS and syphilis, and infectious diseases such as active tuberculosis; or 4) have any circumstances that could affect enrollment.

The Institutional Review Board has approved the study at the First Affiliated Hospital of Xi'an Jiaotong University (Xi'an, China). Investigators explain the research aims and protocol to eligible participants in a face-to-face manner, and those who agree to participate in the BMSC study need to sign the informed consent for participation, for storage of biological samples (including blood, urine, stool, and hair), and for obtaining medical records of both baseline and follow-up surveys.

The BMSC study started in August 2021 and is still ongoing. By August 2024, a total of 1589 participants were recruited. Of them, 951 were men (59.8%), and the mean age was 47 years. The characteristics of the participants are presented in Table 1.

Table 1 Baseline Characteristics of Participants in the BMSC Study Up to August 2024

Characteristics	Total (n = 1589)	Male (n = 951)	Female (n = 638)
Demographic information			
Age, years	46.7 \pm 13.7	48.0 \pm 12.7	44.7 \pm 14.9
Education			
Senior high school and below, n	496 (31.2)	272 (28.6)	224 (35.1)
College and above, n	1093 (68.8)	679 (71.4)	414 (64.9)
Marital status			
Unmarried, n	180 (11.3)	82 (8.6)	98 (15.4)
Married, n	1409 (88.7)	869 (91.4)	540 (84.6)
Profession			
Employed, n	986 (62.1)	644 (67.7)	342 (53.6)
Unemployed, n	603 (38.0)	307 (32.3)	296 (46.4)
Monthly income, ¥			
<3000	351 (22.1)	150 (15.8)	201 (31.5)
\geq 3000	1238 (77.9)	801 (84.2)	437 (68.5)
Census register			
Shaanxi Province, n	1315 (82.8)	803 (84.4)	512 (80.2)
Others, n	274 (17.2)	148 (15.6)	126 (19.8)
Source			
Hospital	1420 (89.4)	861 (90.5)	559 (87.6)
Public areas	169 (10.6)	90 (9.5)	79 (12.4)
Anthropometry			
Body mass index, kg/m ²	27.9 \pm 6.2	27.0 \pm 5.6	29.2 \pm 6.8
Waist circumference, cm	96.9 \pm 14.0	96.7 \pm 12.6	97.1 \pm 16.0
Hip circumference, cm	100.9 \pm 11.1	99.5 \pm 9.7	102.9 \pm 12.7
Biceps circumference, cm	31.3 \pm 4.8	30.8 \pm 4.1	32.0 \pm 5.6
Thigh circumference, cm	53.6 \pm 8.3	52.2 \pm 7.0	55.6 \pm 9.5

(Continued)

Table 1 (Continued).

Characteristics	Total (n = 1589)	Male (n = 951)	Female (n = 638)
Biological samples, n (%)			
Fasting blood	1589 (100.0)	951 (100.0)	638 (100.0)
Urine	1510 (95.0)	914 (96.1)	596(93.4)
Stool	1307 (82.3)	790 (83.1)	517 (81.0)
Hair	1248 (78.5)	632 (66.5)	616 (96.6)
Metabolic disorders, n (%)			
Hyperglycemia	1286 (80.9)	848 (89.2)	438 (68.9)
Obesity	598 (37.6)	283 (29.8)	315 (49.4)
Hypertension	771 (48.5)	490 (51.5)	281 (44.0)
Dyslipidemia	1062 (66.8)	674 (70.9)	388 (60.8)
MetS	1046 (65.8)	653 (68.7)	393 (61.6)
MetS components			
0	13 (0.8)	4 (0.4)	9 (1.4)
1	184 (11.6)	100 (10.5)	84 (13.2)
2	346 (21.8)	194 (20.4)	152 (23.8)
3	425 (26.8)	261 (27.4)	164 (25.7)
4	406 (25.6)	267 (28.1)	139 (21.8)
5	215 (13.5)	125 (13.1)	90 (14.1)

Abbreviations: BP, blood pressure; BMSC, Biobank for Metabolic Syndrome Consequences; FBG, fasting blood glucose; HDL-C, high-density lipoprotein cholesterol; MetS, metabolic syndrome; TG, triglyceride.

Recruitment Strategies

Participants are recruited through two strategies. First, patients presented in the Department of Endocrinology and Metabolism at the First Affiliated Hospital of Xi'an Jiaotong University are evaluated by doctors. The patients who meet the inclusion criteria for the BMSC study are invited and referred to the investigators. Second, participants are recruited through online advertisements and posters posted in public areas.

Study Design

The BMSC study is a prospective observational study of participants aged 18–75 years. After an overnight fast for the baseline survey, eligible participants are invited to the research center or the inpatient ward. Trained researchers conduct the anthropometry according to standard methods. The participants are also asked to complete a structural questionnaire at the first visit to report the basic demographic information, history of disease and medication use, family history of disease, lifestyle and behaviors, physical activity, diet, life events, symptoms of osteoarthritis, sleep, depression, and anxiety. All participants need to provide biological samples according to the study protocol. Professional nurses collect fasted blood, while hair is collected by trained researchers. Urine and stool are collected by patients after the researchers explain the sampling methods. Extensive examinations, such as bioelectrical impedance analysis (BIA), electrocardiogram (ECG), ultrasonography, electromyography (EMG), and fundus photography, are also performed. Echocardiography is performed in a subgroup of patients who need this examination after their physician's evaluation. The measurements are listed in [Table 2](#).

After successful enrollment, follow-up is conducted every 3 months thereafter until the 5-year mark. Anthropometry is performed at every follow-up, and biological samples are collected every two follow-ups. The changes in metabolic indicators and incidence of CMDs and mortality are followed up for at least 5 years ([Figure 1](#)).

Table 2 Overview Measurements within the BMSC Study

Measurements	Content
Questionnaire	Demographic information, socioeconomic status, history of the disease and medication use, disease history of first-degree relatives, smoking behavior, alcohol consumption, physical activity, life events, dietary intake, sleep, symptoms of osteoarthritis, mental health (anxiety and depression), birth weight, and weight change
Anthropometry	Height, weight, waist circumference, hip circumference, abdomen circumference, biceps circumference, and thigh circumference
Biobanking	Fasting serum, EDTA-plasma, RNA, and DNA Spot morning urine samples Stool samples Hair samples
Physical examinations	Blood pressure, heart rate, and area of subcutaneous and visceral fat Bioelectrical impedance analysis, electrocardiogram, electromyography, ultrasonography (neck vessels and lower limbs), and echocardiography

Abbreviation: BMSC, Biobank for Metabolic Syndrome Consequences.

Data Collection

Questionnaires

The participants are asked to complete the structural questionnaire at the baseline survey, including questions on basic characteristics, history of diseases (including but not limited to diabetes mellitus, hypertension, hyperlipidemia, hyperuricemia, polycystic ovary syndrome, coronary heart disease, stroke, congestive heart failure, peripheral arterial disease, cancer, and thyroid disease), and medication use, disease history of first-degree relatives, lifestyle and behaviors, physical activity, life events, dietary intake, sleep, symptoms of osteoarthritis, and mental health (anxiety and depression).

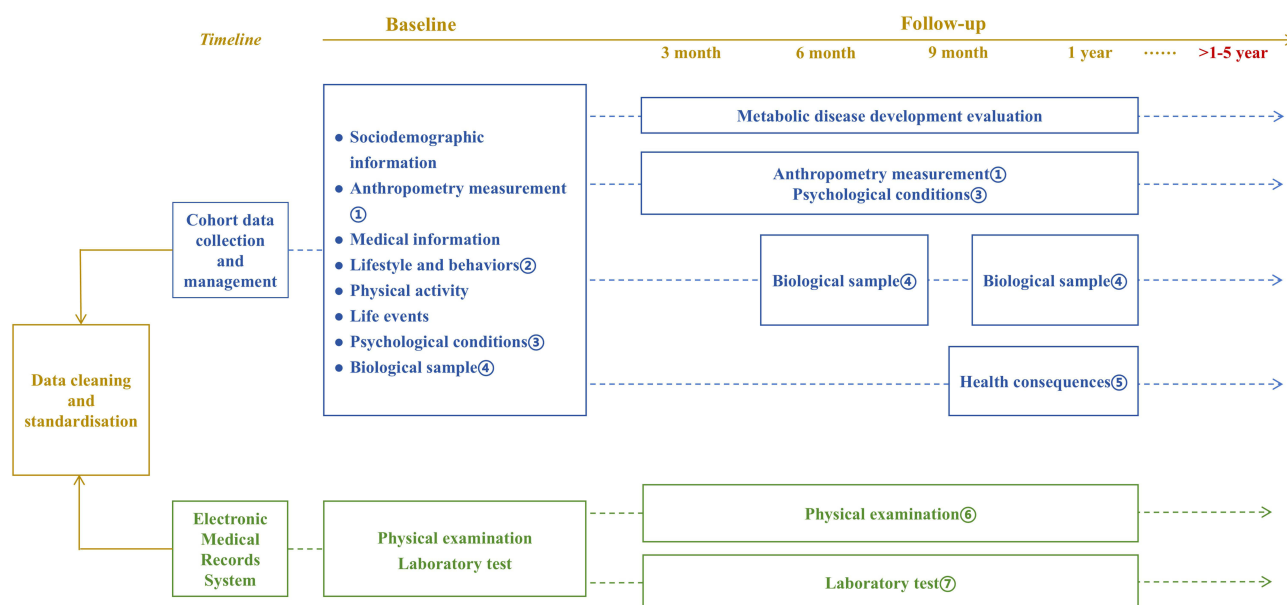


Figure 1 The data framework of the Biobank for Metabolic Syndrome Consequences study. ① Height, weight, WC, BP; ② Smoking, drinking, sleeping, and dietary intake; ③ Depression, anxiety; ④ Blood, urine, stool, and hair; ⑤ CMDs, MetS-related complications; ⑥ Bioelectrical impedance analysis, electrocardiogram, and ultrasonography, etc.; ⑦ Glucose, glycosylated hemoglobin, serum lipid profile, renal function, and liver function, etc. Blue text indicates data actively collected by researchers as part of cohort requirements, whereas green text represents clinical data, with specific tests determined primarily by physicians according to patients' conditions. Yellow and red text denote follow-up time points.

Abbreviations: BP, blood pressure; CMD, cardiometabolic disease; MetS, metabolic syndrome; WC, waist circumference.

Socioeconomic data, including education level, occupation, household income, and insurance status, are collected through standardized, interviewer-administered questionnaires. Details on the requested information are as follows:

Physical Activity

Participants' physical activity is evaluated by the Chinese version of the International Physical Activity Questionnaire (IPAQ)-long form, which has been well-validated in the Chinese population.^{23,24}

Dietary Intake

The participants report a long-term dietary intake through a semi-quantitative food frequency questionnaire (FFQ). Because the FFQ is relatively easy to use, inexpensive, and reflects a long-term dietary intake pattern, it has become the most common tool to evaluate the diet in large epidemiological studies. The FFQ used in the BMSC study covers eighteen categories: grain (11 items); beans (8 items); fresh beans (5 items); root vegetable (5 items); melon vegetable (7 items); leaf vegetable (16 items); fruit (4 items); nut (4 items); animal meat (13 items); poultry (5 items); milk (1 item); eggs (4 items); aquatic products (8 items); fungi foods (5 items); pickle (4 items); alcohol and other beverages (5 items); oil (6 items); condiments (8 items); and nutritious supplementation.

Sleep

Sleep assessment is performed via three questionnaires. Epworth sleepiness scale (ESS) is used to assess daytime sleepiness,²⁵ the Pittsburgh sleep quality index (PSQI) is used to evaluate the quality of sleep,²⁶ and the Berlin questionnaire is used for assessment of the presence of obstructive sleep apnea syndrome.²⁷

Symptoms of Osteoarthritis

Symptoms of osteoarthritis are scaled by the Knee Injury and Osteoarthritis Outcome Score (KOOS), which is developed to assess the patient's opinion on knee health and related problems and has been validated in the Chinese population.²⁸

Mental Health

Both depression and anxiety are evaluated. The Zung Self-rating Depression Scale (SDS) is used to determine depressive symptoms via twenty terms. Indicators of anxiety are collected using the Beck Anxiety Inventory (BAI). SDS²⁹ and BAI³⁰ have been well validated in the Chinese population.

Physical Examination

All participants undergo a physical examination at baseline survey, including anthropometry measurement, blood pressure, heart rate, subcutaneous fat, and visceral fat. Height and weight are measured using an ultrasonic height and weight meter (OMRON MEDICAL Beijing Co., Ltd. HNH-318) to the nearest 0.01 m and 0.1 kg, respectively. The waist circumference is the circumference of the midpoint line between the lowest point of the rib and the upper margin of the iliac crest (the narrowest point of the waist) measured at the end of exhalation and before the beginning of inhalation using a tape measure, with a precision of 0.1 cm. Hip circumference is measured as the maximum circumference of the buttocks. Abdomen circumference is measured at the iliac crest point at the end of exhalation but before the beginning of inhalation. The biceps circumference is measured at the upper 1/3 of the right arm, while the thigh circumference is measured at the upper 1/3 of the right thigh. An electronic blood pressure monitor (OMRON MEDICAL Beijing Co., Ltd. HEM-7121) is used to measure the blood pressure and heart rate, after the participants are asked to be quiet and rest for at least 5 minutes. The subcutaneous and visceral fat area is measured using a visceral fat detector (OMRON MEDICAL Beijing Co., Ltd. 3 DUALSCAN, HDS-2000), with a unit of cm².

Advanced examinations such as BIA, ECG, and ultrasonography are also performed. The body composition, including body fat (body fat percentage), muscle mass (skeletal muscle content), and visceral fat (visceral fat grade), is measured using the direct segmental multifrequency BIA method DSM-BIA (InBody H20) to the nearest 0.1 kg. A resting 12-lead ECG (Mortara Eli-350) is recorded and archived electronically at baseline survey. Carotid intima-media thickness is measured by color Doppler ultrasonography of the neck vessels. Peripheral arterial disease is determined with the help of color Doppler ultrasonography of lower limbs. Fundus photography of both eyes is performed to determine the presence of diabetic retinopathy. Surface EMG assessment is only performed to assist in the diagnosis of

diabetic peripheral neuropathy. Echocardiography is performed in a subset of individuals needing examinations after their physicians' evaluation.

Fasting Blood Sampling

A total of 20 mL peripheral venous blood is taken after an overnight fast of at least 8 hours, and used for the collection of serum, EDTA-plasma, RNA, and DNA, all stored in a -80°C freezer for future multi-omics analyses. Separate blood samples are taken for immediate tests, such as glucose, HbA1c levels, serum lipid profile (total cholesterol, triglycerides, HDL-C, LDL-C, apolipoprotein A1, total apolipoprotein B, and apolipoprotein E), indicators of liver function (aspartate transaminase, alanine transaminase, total protein, albumin, globulin, total bile acids, alkaline phosphatase, γ -glutamyltranspeptidase, total bilirubin, direct bilirubin, and indirect bilirubin), indicators of renal function (urea, cystatin C, uric acid, creatinine, glycated albumin, and retinol-binding protein), indicators of thyroid function (serum thyroid stimulating hormone, total and free thyroxine, total and free triiodothyronine), and blood cell count, mean red blood cell (RBC) volume, platelet concentration, platelet distribution width, mean platelet volume, and high-sensitivity C-reactive protein (CRP). Insulin release test, C-peptide release test, and serum insulin and C-peptide measurements are only performed in a subset of individuals who need these examinations according to the disease.

Urine Sampling

A tube of 15 mL urine is obtained from all participants at enrollment. The participants are asked to fast for at least 8 hours before the collection of midstream urine. Collected urine samples are stored in a -80°C freezer for future analyses. The urine samples used for routine analyses are separately collected. Creatinine and urinary microalbumin are tested using standard methods at the clinical laboratory.

Stool Sampling

After urination, two tubes of fecal samples are collected. After receiving the fecal samples, the researchers put them into liquid nitrogen for rapid freezing and then store them at -80°C freezer for future analyses.

Hair Sampling

Using blunt curved scissors, the researchers collect 30 to 50 hairs from the occipital bone, as close as possible to the root of the hair, taking 3 cm. The hair sample is rolled in tinfoil and stored at room temperature.

Ascertainment of Incident Diseases During Follow-Up

The primary endpoints of the BMSC study include major CMDs, such as CVDs (coronary heart disease, stroke, myocardial infarction, and congestive heart failure), MASLD, mortality, and other MetS-related diseases (T2D, depression, and osteoarthritis). Endpoints are ascertained through three complementary sources: (1) standardized questionnaires administered annually, including self-reported physician diagnoses; (2) biochemical indicators collected every six months, with prespecified thresholds (e.g., fasting glucose, HbA1c, lipid profile, liver function tests) used to define disease onset or progression; and (3) electronic medical records from the First Affiliated Hospital of Xi'an Jiatong University, linked and extracted two years after enrollment and subsequently at 1-year intervals. Diagnoses from medical records will be verified using ICD-10 codes and clinical criteria, and, where necessary, adjudicated by an independent panel of physicians to ensure accuracy. This multi-source and standardized approach is designed to enhance both the validity and reliability of endpoint determination, thereby improving the robustness of event-based analysis.

Biological Sample Processing Methods

As mentioned above, common clinical biomarkers, such as blood type and counts, glucose, HbA1c, lipids, indicators of liver function, kidney function, and thyroid function, are tested using automatic equipment in the hospital laboratory (Table 3). Other advanced assays are performed as follows:

Blood samples are used for genomics according to the standard protocols. After DNA is extracted from white cells, the samples are delivered to a professional sequencing company for high-throughput genotyping after strict quality control such as gel electrophoresis, NanoDrop, and Bioanalyzer (Infinium CoreExome-24 BeadChip is intended to be selected).

Table 3 Laboratory Analyses at Baseline within the BMSC Study

Laboratory Analyses	Analysis Method
Serum/plasma concentrations	
Alanine transaminase	Velocity assays
Albumin	Bromocresol green method
Alkaline phosphatase	AMP buffer method
Apolipoprotein A1, apolipoprotein B, apolipoprotein E	Immunoturbidimetric
Aspartate transaminase	Velocity assays
Blood cell count	Flow cytometry method
Concentration of platelet	Impedance method
Creatinine	Enzymatic
Cystatin C	Immunoturbidimetric
Direct bilirubin	Vanadate oxidation method
Free thyroxine	Chemiluminescence
Free triiodothyronine	Chemiluminescence
Globulin	Calculation with total protein and albumin
Glucose	Hexokinase method
Glycosylated hemoglobin	HPLC
Hemoglobin	Colorimetric analysis
High-density lipoprotein cholesterol	Selective inhibition method
High-sensitivity C-reactive protein	Immunoturbidimetric
Indirect bilirubin	Calculation with total bilirubin and direct bilirubin
Insulin	Direct chemiluminescence
Low-density lipoprotein cholesterol	Selective direct method
Mean platelet volume	Impedance method
Mean red cell volume	Calculation
Platelet distribution width	Calculation
Prealbumin	Immunoturbidimetric
Retinol-binding protein	Latex immunoturbidimetric assay
Thyroid-stimulating hormone	Chemiluminescence
Thyroxine	Chemiluminescence
Total bile acid	Enzymatic cycling method
Total bilirubin	Vanadate oxidation method
Total cholesterol	Cholesterol oxidase
Total protein	Biuret method
Triglyceride	Deglycerin method
Triiodothyronine	Chemiluminescence
Urea	Urease UV kinetic assay
Uric acid	Enzymatic
γ -glutamyltranspeptidase	Velocity assays
Urine (morning spot)	
Albumin	Chemical analysis methods
Creatinine	Enzymatic

Abbreviations: BMSC, Biobank for Metabolic Syndrome Consequences; HPLC, high-performance liquid chromatography.

Plasma concentrations of cholesterol and its precursors, including desmosterol, are measured using rapid gas chromatogramtography-mass spectrometry (GC-MS). Plasma inflammatory factors such as ICAM-1, TNF- α , IL-6, and TGF- β are assessed by ELISA kits.

Fecal samples are mainly intended for microbiome analysis, according to established protocols. Genomic DNA is extracted from fecal samples using the QIAamp DNA Stool Mini Kit (Qiagen, Germany) according to the manufacturer’s instructions and then used for 16S rRNA gene sequencing.

Plasma and fecal samples are also intended for assessing metabolomics and lipidomics. Levels of steroid hormones in hair samples are determined through liquid chromatography-mass spectrometry (LC-MS).

Quality Control

Several strategies are conducted to ensure the quality of the BMSC study (Figure 2). First, the BMSC study is led by a professional data collection team, and every data collector must have attended a professional training program. Additionally, a handbook for data collection has been developed to facilitate the daily work. Second, both artificial and system queries are performed to ensure the quality of the investigation and collection of anthropometry data. Third, the biological sample should be sent to the Biobank as soon as possible to ensure the quality of the samples after they have been withdrawn. Last, in the laboratory where the analysis of samples takes place, researchers should operate with strict experimental methods to ensure the validity of experimental data.

Statistical Analysis

The BMSC study encompasses multiple research aims and involves a wide range of exposures and outcomes. To ensure adequate statistical power, we performed a sample size estimation using T2D as the exposure and stroke as the outcome. Based on results from the Jinchang Cohort Study in Gansu Province, the incidence of stroke was estimated to be 5.26% among participants with T2D and 1.71% among those without T2D. Assuming a two-sided significance level (α) of 0.05, statistical power ($1-\beta$) of 80%, and a 1:1 matching ratio, the required sample size to detect a significant difference between groups was calculated to be 840. To account for potential loss to follow-up and withdrawal during long-term observation, we increased the sample size by 20%, resulting in 1008 participants. Considering that the study aims to investigate multiple outcomes beyond stroke, we determined that a baseline sample size of approximately 2000 participants would be appropriate.

Appropriate statistical methods will be used to analyze and present the basic characteristics. Regression models (e.g., logistic regression model) will be fitted to analyze the cross-sectional associations between influence factors and metabolic disorders, while potential prospective associations between endpoint diseases and influence factors will be assessed by linear mixed-effects models combined with Cox regression. Since BMSC is an open, prospective cohort, the baseline dates differ among participants based on their recruitment date. The variation in follow-up duration is fully

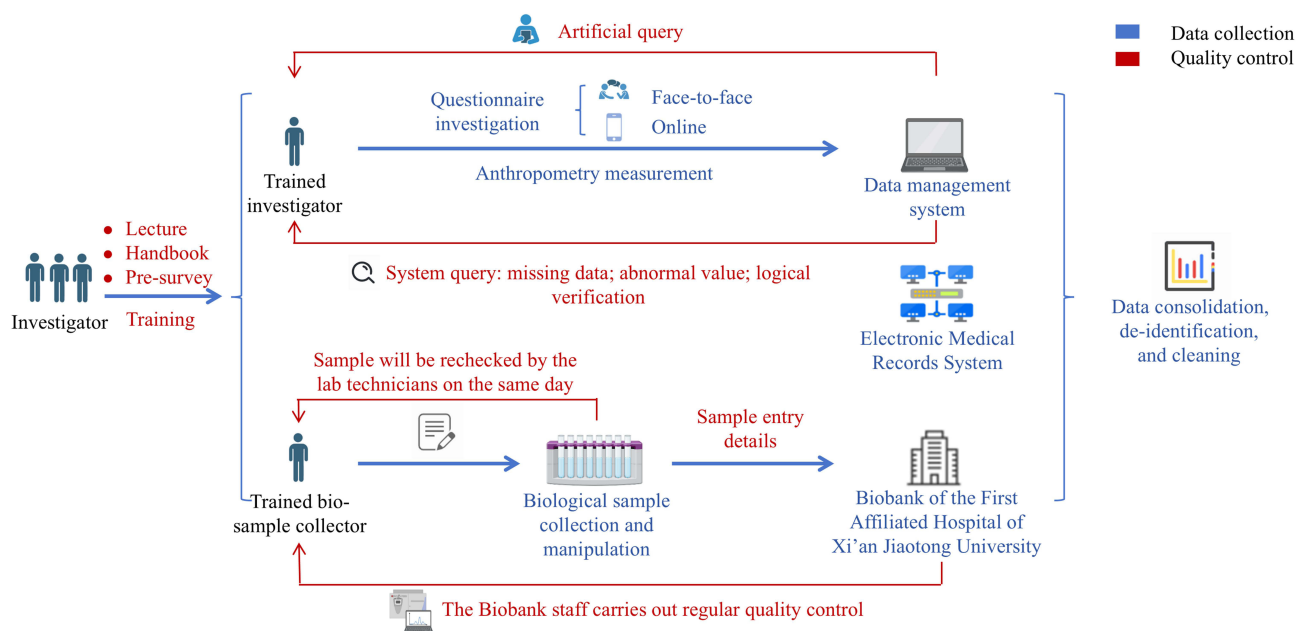


Figure 2 The process of data collection and quality control of the Biobank for Metabolic Syndrome Consequences study. The blue texts and arrows represent the data and sample collection process, while the red color indicates the quality control process during data collection.

acknowledged and will be addressed through time-to-event analysis. Most core covariates (age, sex, anthropometrics) are fully recorded at baseline. For questionnaire-derived data (e.g., lifestyle, socioeconomic status), missing covariates will be handled via multiple imputation where appropriate.

For microbiome measurements, the Kruskal–Wallis test or regression analysis will be conducted to compare the α -diversity index, such as the Shannon index, Observed feature, and the Gini-Simpson Index. Species-level analysis will be performed using MaAsLin or multiple regression.

For metabolomics and lipidomics data, the least absolute shrinkage and selection operator and principal component analysis will be used to reduce the dimensionality of the data. Correlation analysis, regression analysis, and machine learning methods (e.g., Light Gradient Boosting Machine) will be applied to identify associations between omics features and clinical outcomes. To integrate multi-omics datasets (genomics, metabolomics, microbiome) with epidemiological and clinical variables, we will employ data fusion approaches such as multi-view learning and network-based methods, enabling the identification of cross-omics biomarkers and molecular signatures. Pathway enrichment analysis will further aid in interpreting the biological relevance. This integrative framework is expected to support biomarker discovery, risk prediction, and provide mechanistic insights into the pathophysiology of MetS and its consequences.

Basic Characteristics of the Participants at Recruitment

Up to August 2024, a total of 1589 participants have been included in the BMSC study. Of them, 951 are men (59.8%) and 1420 (89.4%) are enrolled in the hospital. The age of participants ranged from 18 to 74 [mean = 47, standard deviation (SD): 14]. The average BMI was 27.0 kg/m² (SD: 5.6) for men and 29.2 kg/m² (SD: 6.9) for women. The total rate of biological sample collection was at least 78.5%. Hyperglycemia was the most prevalent metabolic disorder in the participants (80.9%). 37.6% of participants were obese, and the prevalence of MetS was 65.8%. Among the study population, the proportions of participants with 0, 1, 2, 3, 4, and 5 MetS components were 0.8%, 11.6%, 21.8%, 26.8%, 25.6%, and 13.5%, respectively (Table 1). Table 4 presents the distribution of several cardiometabolic biomarkers. Fasting plasma glucose, HbA1c, TyG, triglycerides, systolic blood pressure, diastolic blood pressure, and serum uric acid were higher in males, while total cholesterol, HDL-C, and LDL-C were higher in females (all *P* values \leq 0.026).

Strengths and Limitations

The BMSC study is an ongoing prospective observational study of approximately 2000 adults with metabolic disorders. It has several strengths. First, the study's deep phenotyping yields profiles of individual characteristics in many dimensions, including anthropometry, lifestyle factors, psychosocial characteristics, medical records, clinical biomarkers, and multi-omics profiles. Second, strict measures adopted during the project ensure high-quality data collection. Third,

Table 4 Cardiometabolic Biomarkers of Study Participants in the BMSC Study

Cardiometabolic Biomarkers	Total (n = 1589)	Male (n = 951)	Female (n = 638)	<i>P</i> value*
FPG, mean (SD), mmol/L	7.7±3.5	8.2±3.9	7.1±2.8	<0.001
HbA1c, mean (SD), %	8.3±2.0	8.6±2.1	7.8±1.7	<0.001
TyG	9.1±0.8	9.2±0.8	9.0±0.7	<0.001
TC, mean (SD), mmol/L	4.4±1.1	4.3±1.2	4.5±1.0	<0.001
TG, mean (SD), mmol/L	2.0±1.8	2.1±2.0	1.9±1.4	0.019
HDL-C, mean (SD), mmol/L	1.0±0.2	1.0±0.2	1.1±0.3	<0.001
LDL-C, mean (SD), mmol/L	2.6±0.8	2.5±0.9	2.7±0.8	<0.001
SBP, mean (SD), mm Hg	129.6±16.3	130.4±16.4	128.4±16.2	0.019
DBP, mean (SD), mm Hg	80.0±11.2	81.1±11.1	78.3±11.1	<0.001
SUA, mean (SD), mmol/L	355.8±93.8	360.1±98.4	349.4±86.0	0.026

Notes: **P* values were calculated (between males and females) by two-sample Student's *t* test.

Abbreviations: BMSC, Biobank for Metabolic Syndrome Consequences; DBP, diastolic blood pressure; FPG, fasting plasma glucose; HbA1c, glycated haemoglobin; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; TC, total cholesterol; TG, triglycerides; SUA, serum uric acid.

the standardized Biobank procedure guarantees the quality of biological samples, which is one of the distinctive strengths of the BMSC study. Fourth, repeated measurements and biological sampling across multiple time points make it possible to assess the dynamic change of influencing factors and biomarkers. However, some limitations should be acknowledged. First, participants in the BMSC study are recruited in Northwest China; hence, extrapolation of results to other regions or countries requires caution. Second, only participants with metabolic disorders meet the inclusion criteria of the BMSC study, which precludes the recruitment of healthy participants as controls. The absence of a control group limits the ability to determine whether the observed alterations are disease-specific or reflect normal variations, thereby reducing the capacity to infer causality and restricting the generalizability of our findings to broader populations. To address this limitation, we have initiated the recruitment of metabolically healthy individuals from the Department of Health Medicine of our hospital, who may serve as an appropriate control group in future analyses. This will allow us to better delineate disease-specific alterations and strengthen the robustness of our conclusions. Third, while observational cohort designs inherently limit causal inference, the BMSC study aims to elucidate potential mechanistic pathways and identify clinically relevant associations, rather than assert definitive causal relationships. In addition, the current sample size is modest compared to national mega-cohorts in China. However, the BMSC study especially focuses on deeply phenotyped, metabolically characterized individuals, with integrated multi-omics datasets and long-term biobanked biospecimens. Finally, several practical challenges should also be acknowledged. First, maintaining participant retention over a 5-year follow-up may be difficult, and loss to follow-up could reduce statistical power. Second, the management and analysis of high-dimensional multi-omics data are complex and will require rigorous bioinformatics pipelines and quality control procedures. Third, as the study is conducted in a single regional population, the generalizability of findings to other geographic or ethnic groups may be limited. These challenges will be carefully considered in the implementation and interpretation of the BMSC study.

Prospective Contributions of the BMSC Study

Although no study results are yet available, the BMSC protocol has several distinctive features that may provide novel insights into the etiology and progression of MetS and related chronic diseases. The repeated biospecimen collection combined with multi-omics profiling and clinical data enables comprehensive characterization of biological pathways. Moreover, the relatively high-frequency follow-up (every 3 months) enhances the granularity of longitudinal assessments, which is uncommon in large cohort studies. Together, these features make the BMSC study a valuable resource with strong potential to contribute to precision medicine research and to inform prevention and intervention strategies for metabolic and cardiovascular diseases. Beyond scientific discovery, the BMSC study may also have important implications for public health and clinical practice. For example, longitudinal biomarker profiling could help to refine risk stratification tools for MetS and related CMDs, thereby informing clinical guidelines for earlier detection and more personalized prevention strategies. In addition, the integration of multi-omics data with lifestyle and clinical outcomes may identify novel therapeutic targets, which could ultimately support the development of precision interventions for metabolic disorders. These potential applications highlight the translational value of the BMSC study beyond research contexts.

Conclusion

The BMSC study provides a new framework for a prospective MetS biobank with multi-dimensional information in an underrepresented population. This resource is expected to contribute to disease prediction, early diagnosis, and mechanistic understanding, thereby informing precision medicine and public health strategies. While we recognize practical challenges such as long-term participant retention, management of complex data, and limited generalizability beyond the study region, these can be addressed through rigorous methodology, robust infrastructure, and collaboration efforts. Continued stakeholder support and open opportunities for data sharing and collaboration will be critical to maximizing the value of the BMSC and ensuring its long-term impact.

Abbreviations

BAI, Beck Anxiety Inventory; BIA, Bioelectrical impedance analysis; BMSC, Biobank for Metabolic Syndrome Consequences; CKD, Chronic kidney disease; CMD, Cardiometabolic diseases; COPD, Chronic obstructive pulmonary disease; CRP, C-reactive protein; CVD, Cardiovascular disease; ECG, Electrocardiogram; EMG, Electromyography; ESS, Epworth sleepiness scale; FFQ, Food frequency questionnaire; HbA1c, Glycosylated hemoglobin; HDL-C, High-density lipoprotein cholesterol; IPAQ, International Physical Activity Questionnaire; KOOS, Knee Injury and Osteoarthritis Outcome Score; LDL-C, Low-density lipoprotein cholesterol; MASH, Metabolic dysfunction-associated steatohepatitis; MASLD, Metabolic dysfunction-associated steatotic liver disease; MetS, Metabolic syndrome; OGTT, Oral glucose tolerance test; PSQI, Pittsburgh sleep quality index; RBC, Red blood cell; SDS, Self-rating Depression Scale; T2D, Type 2 diabetes.

Data Sharing Statement

All data used in this study are from the BMSC study and are available from the corresponding author (Prof. Yanan Wang) upon reasonable request.

Ethical Approval and Consent to Participate

This is an observational study. The Institutional Review Board of the First Affiliated Hospital of Xi'an Jiaotong University (Xi'an, China) has approved the study, and the institutional review board (IRB) reference number is No. XJTU1AF2021LSK-193. All participants provided informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Acknowledgments

The BMSC study is supported by the Department of Endocrinology and Metabolism, Department of Hepatological Survey, and Biobank of First Affiliated Hospital of Xi'an Jiaotong University. We thank all contributing individuals who participate in the BMSC study. We also express our gratitude to all doctors who invite eligible participants.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The study is sponsored by the National Science and Technology Major Project of the Ministry of Science and Technology of China (No. 2024ZD0531702), the Fundamental Research Funds for the Central Universities (No. xtr052023011), the Plan for Strengthening Basic Disciplines of Xi'an (No. 21YXYJ0111), the China Postdoctoral Science Foundation (2024M762615), the National Natural Science Foundation of China (82170876), and the Key Research and Development Program of Shaanxi Province (No. 2025SF-YBXM-367).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Eckel RH, Alberti KG, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet*. 2010;375:181–183. doi:10.1016/S0140-6736(09)61794-3
2. Noubiap JJ, Nansseu JR, Lontchi-Yimagou E, et al. Geographic distribution of metabolic syndrome and its components in the general adult population: a meta-analysis of global data from 28 million individuals. *Diabet Res Clin Pract*. 2022;188:109924. doi:10.1016/j.diabres.2022.109924
3. Huang J, Huang AL, Withers M, et al. Prevalence of metabolic syndrome in Chinese women and men: a systematic review and meta-analysis of data from 734511 individuals. *Lancet*. 2018;392:S14.
4. Huang Y, Zhang L, Wang Z, et al. The prevalence and characteristics of metabolic syndrome according to different definitions in China: a nationwide cross-sectional study, 2012–2015. *BMC Public Health*. 2022;22:1869. doi:10.1186/s12889-022-14263-w

5. Wilson PW, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation*. 2005;112:3066–3072. doi:10.1161/CIRCULATIONAHA.105.539528
6. Shin JA, Lee JH, Lim SY, et al. Metabolic syndrome as a predictor of type 2 diabetes, and its clinical interpretations and usefulness. *J Diabetes Investig*. 2013;4:334–343. doi:10.1111/jdi.12075
7. Paschos P, Paletas K. Non alcoholic fatty liver disease and metabolic syndrome. *Hippokratia*. 2009;13:9–19.
8. Pammer LM, Lamina C, Schultheiss UT, et al. Association of the metabolic syndrome with mortality and major adverse cardiac events: a large chronic kidney disease cohort. *J Intern Med*. 2021;290:1219–1232. doi:10.1111/joim.13355
9. Malik S, Wong ND, Franklin SS, et al. Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults. *Circulation*. 2004;110:1245–1250. doi:10.1161/01.CIR.0000140677.20606.0E
10. Iseki K, Konta T, Asahi K, et al. Impact of metabolic syndrome on the mortality rate among participants in a specific health check and guidance program in Japan. *Intern Med*. 2020;59:2671–2678. doi:10.2169/internalmedicine.4975-20
11. Scurt FG, Ganz MJ, Herzog C, Bose K, Mertens PR, Chatzikyrkou C. Association of metabolic syndrome and chronic kidney disease. *Obes Rev*. 2024;25:e13649. doi:10.1111/obr.13649
12. Stefansson VTN, Schei J, Solbu MD, Jenssen TG, Melsom T, Eriksen BO. Metabolic syndrome but not obesity measures are risk factors for accelerated age-related glomerular filtration rate decline in the general population. *Kidney Int*. 2018;93:1183–1190. doi:10.1016/j.kint.2017.11.012
13. Chan SMH, Selemidis S, Bozinovski S, Vlahos R. Pathobiological mechanisms underlying metabolic syndrome (MetS) in chronic obstructive pulmonary disease (COPD): clinical significance and therapeutic strategies. *Pharmacol Ther*. 2019;198:160–188. doi:10.1016/j.pharmthera.2019.02.013
14. Li S, Zhang T, Yang H, et al. Metabolic syndrome, genetic susceptibility, and risk of chronic obstructive pulmonary disease: the UK Biobank Study. *Diabetes Obes Metab*. 2023;26(2):482–494. doi:10.1111/dom.15334
15. Pan A, Keum N, Okereke OI, et al. Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies. *Diabetes Care*. 2012;35:1171–1180. doi:10.2337/dc11-2055
16. Wang X, Ji L, Tang Z, et al. The association of metabolic syndrome and cognitive impairment in Jidong of China: a cross-sectional study. *BMC Endocr Disord*. 2021;21:40. doi:10.1186/s12902-021-00705-w
17. Codazzi V, Frontino G, Galimberti L, Giustina A, Petrelli A. Mechanisms and risk factors of metabolic syndrome in children and adolescents. *Endocrine*. 2024;84:16–28. doi:10.1007/s12020-023-03642-x
18. Dang AK, Le HT, Nguyen GT, et al. Prevalence of metabolic syndrome and its related factors among Vietnamese people: a systematic review and meta-analysis. *Diabetol Metab Syndr*. 2022;16:102477. doi:10.1016/j.dsx.2022.102477
19. Matsuzawa Y, Funahashi T, Nakamura T. The concept of metabolic syndrome: contribution of visceral fat accumulation and its molecular mechanism. *J Atheroscler Thromb*. 2011;18:629–639. doi:10.5551/jat.7922
20. Reddy P, Lent-Schochet D, Ramakrishnan N, McLaughlin M, Jialal I. Metabolic syndrome is an inflammatory disorder: a conspiracy between adipose tissue and phagocytes. *Clin Chim Acta*. 2019;496:35–44. doi:10.1016/j.cca.2019.06.019
21. Association AD. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2021. *Diabetes Care*. 2021;44:S15–S33. doi:10.2337/dc21-S002
22. Joint Committee Issued Chinese Guideline for the Management of Dyslipidemia in Adults. [2016 Chinese guideline for the management of dyslipidemia in adults]. *Zhonghua xin xue guan bing za zhi*. 2016;44:833–853. Polish. doi:10.3760/cma.j.issn.0253-3758.2016.10.005
23. Macfarlane D, Chan A, Cerin E. Examining the validity and reliability of the Chinese version of the International Physical Activity Questionnaire, long form (IPAQ-LC). *Public Health Nutr*. 2011;14:443–450. doi:10.1017/S1368980010002806
24. Jia YJ, Xu LZ, Kang DY, Tang Y. Reliability and validity regarding the Chinese version of the International Physical Activity Questionnaires (long self-administrated format) on women in Chengdu, China. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2008;29:1078–1082.
25. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14:540–545. doi:10.1093/sleep/14.6.540
26. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28:193–213. doi:10.1016/0165-1781(89)90047-4
27. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann Intern Med*. 1999;131:485–491. doi:10.7326/0003-4819-131-7-199910050-00002
28. Cheung RT, Ngai SP, Ho KK. Chinese adaptation and validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in patients with knee osteoarthritis. *Rheumatol Int*. 2016;36:1449–1454. doi:10.1007/s00296-016-3539-7
29. Lee HC, Chiu HF, Wing YK, Leung CM, Kwong PK, Chung DW. The Zung self-rating depression scale: screening for depression among the Hong Kong Chinese elderly. *J Geriatr Psychiatry Neurol*. 1994;7:216–220. doi:10.1177/089198879400700404
30. Che -H-H, Lu M-L, Chen H-C, Chang S-W, Lee Y-J. Validation of the Chinese version of the Beck Anxiety Inventory. *Formos J Med*. 2006;10:451–452.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress
Taylor & Francis Group