BMJ Open

Cohort profile: the Liyang cohort study on chronic diseases and risk factors monitoring in China (Liyang Study)

Liang Zhou,[1] Wei Hu ,[2] Siyuan Liu,[2] Yanan Qiao,[2] Dingliu He,[2] Shuting Xiong,[2] Liuming Peng,[1] Lei Cao,[1] Ying Wu,[1] Na Sun,[2] Qiang Han,[2] Jiadong Chu,[2] Xuanli Chen,[2] Tongxing Li,[2] Zhaolong Feng,[2] Qida He,[2] Chaofu Ke ,[2] Yueping Shen[2]

To cite: Zhou L, Hu W, Liu S, *et al*. Cohort profile: the Liyang cohort study on chronic diseases and risk factors monitoring in China (Liyang Study). *BMJ Open* 2022;**12**:e060978. doi:10.1136/bmjopen-2022-060978

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-060978).

LZ and WH contributed equally.

LZ and WH are joint first authors.

Received 11 January 2022
Accepted 27 June 2022

[1]Liyang Center for Disease Control and Prevention, Liyang Health Bureau, Liyang, China
[2]Department of Epidemiology and Biostatistics, Soochow University Medical College, Suzhou, China

Correspondence to
Dr Yueping Shen; shenyueping@suda.edu.cn

ABSTRACTS

Purpose The Liyang cohort study on chronic diseases and risk factors monitoring in China (Liyang Study) is a prospective population-based study which aims to investigate and identify the determinants of the most prevalent chronic non-communicable diseases (NCDs) and to evaluate the impact of demographic characteristics, lifestyle, dietary habits, cognition, disability and NCDs on the health-related quality of life.

Participants Between March 2019 and June 2020, 10 056 individuals aged ≥18 years were administered a baseline survey through a multistage cluster random sampling in Liyang City, southern Jiangsu Province, China.

Findings to date The Liyang Study included detailed sociodemographic, anthropometric and health-related behaviour, common NCDs and blood sample information. Moreover, the study gathered a series of data on specific scales including the activities of daily living, instrumental activities of daily living, abbreviated mental test, Food Frequency Questionnaire and EuroQol 5-Dimensions 5-Levels Scale. Of the 10 056 participants, 52.92% (n=5322) were female and 92.26% (n=9278) came from rural areas. The mean age was 49.9±16.2 years. Men were more likely to have a higher level of education, annual income and a paid job than women ($p<0.05$). The top three overall most prevalent NCDs in the study were hypertension (18.06%, n=1815), digestive diseases (7.88%, n=791), and arthritis or rheumatism (5.28%, n=530). Women had a significantly higher prevalence of diabetes (5.46%, n=290 vs 4.42%, n=209, p=0.016) and arthritis (6.04%, n=321 vs 4.42%, n=209, $p<0.001$) than men, while the opposite was true for chronic lung diseases such as chronic obstructive pulmonary disease (1.37%, n=65 vs 0.92%, n=49, p=0.032) and chronic hepatic diseases (0.80%, n=38 vs 0.47%, n=25, p=0.035).

Future plans The current study will give valuable insights into the association between sociodemographic factors, health-related behaviour, diet, cognition, disability and genetic factors and the most prevalent NCDs among local community residents. Starting from 2022, a follow-up survey will be conducted every 3 years to further explore the causal relationship between the above factors and NCDs.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The current study provides reliable representative data from the entire population, based on a large sample size (n=10 056) and a wide range of variables, providing substantial evidence for identifying and measuring the prevalence and risk factors of common NCDs.
⇒ The collection of several biological samples at baseline, including whole blood, plasma, serum and white blood cell samples, will help effectively capture and determine relevant information about the genetics and variation of NCDs.
⇒ Extensive and detailed baseline data included not only basic sociodemographic and lifestyle information but also detailed scale data (including Food Frequency Questionnaire, abbreviated mental test, activities of daily living, instrumental activities of daily living and EuroQol 5-Dimensions 5-Levels).
⇒ Self-reported data from NCDs (especially mental and memory-related disorders) may produce information biases, such as recall bias.
⇒ The under-representation of participants with urban residence status in this cohort may lead to bias in estimating the association of NCDs among the general population in Liyang city.

INTRODUCTION

Owing to remarkable changes in lifestyle, severe environmental pollution and the ageing population in China, chronic non-communicable diseases (NCDs) have become a major Chinese public health problem, with approximately 260 million NCD cases recently diagnosed.[1–3] According to the 2013 National Health Services Survey, the prevalence of NCDs in the general population has reached 24.5%, and the direct financial burden of NCDs is estimated to climb to $21.72 billion in China.[4] Unfortunately, even though significant efforts have been made to discover novel biomarkers and develop targeted strategies for preventing NCDs, the incidence and prevalence rates of NCDs will continue to increase in the coming years.[5] Moreover, most regions of China face a unique dilemma regarding the imbalance of health resources between rural and urban areas, as well as the significant burden of medical insurance systems, which may further exacerbate the risks of functional impairments and life-threatening events due to NCDs. Given the growing burden and high prevalence and mortality rates of NCDs, it is important to identify new potential risk factors to develop more effective and targeted intervention measures, which may guarantee optimal utilisation of medical resources, especially in resource-unbalanced regions.

NCDs place a heavy burden on healthcare systems, both financially and structurally, because of their long duration and slow progress.[6] Due to the uneven coverage of medical service resources and healthcare systems in resource-imbalanced regions, premature death, disability and medical expenditure of patients with NCDs are more pronounced.[4] Besides, the proportion of population health risk factors continues to grow, and the high differentiation between medical services and health management services leads to a break point in the chain of systematic diagnosis, treatment and management services of NCDs.[7] These adverse factors will lead to greater pressure on the prevention and control of NCDs in resource-imbalanced regions.[6]

The design of community-based prospective studies and the follow-up lasting several years are important in establishing associations between slow-acting causes and particular diseases. In the 1990s, a large-scale longitudinal study based on 45 nationally representative disease surveillance points was conducted to investigate a series of easily modified determinants of NCDs in China.[8] This study was one of the earlier studies that revealed the health effects of tobacco consumption, alcohol intake, body mass index (BMI) and blood pressure on NCDs. However, the study did not collect blood samples, thus offering no evaluation of blood-related risk factors, such as genetic variants, microRNA and metabolites. To address this gap, a blood-based prospective study involving 500 000 middle-aged adults from 10 different regions of China has been conducted since 2004.[9] However, this study did not measure the quality of life of the participants. It is worth noting that people with NCDs generally have a poor quality of life.[10] Consequently, the quality of life and its determinants in the general population warrant further research, especially in those afflicted by NCDs. Still, researchers will be interfered with by time-dependent confounding factors when searching for the exact cause of NCDs.[11] Hence, it is imperative to establish a cohort study that meets the following requirements. First, like the existing NCD cohort studies,[9 12] the risk factors and prevalence of NCDs should be explored. Second, a new indicator of health outcomes, health-related quality of life (HRQoL) indicator, should be included. Studies have shown that in addition to the objective indicators such as morbidity, mortality and clinical measurement, HRQoL (the physiological, psychological and social well-being reported by the patient) has become an important part of the evaluation and monitoring of NCDs.[3 10] Finally, a cohort study should provide a novel source of data covering the entire age group to develop new statistical methods of causal inference that also take into account time-dependent confounders.

Liyang is located in the south of Jiangsu Province, with a permanent, year-round population of about 0.76 million in 2020. As an economically developed city at the crossroads of the Nanjing metropolitan area and Shanghai economic zone, the epidemic situation of NCDs in Liyang city could provide a reference for most parts of the developed areas of China. The Liyang Study was launched in Liyang from March 2019 to June 2020 based on a standardised survey procedure and strict quality control process. The purposes of this cohort study are fivefold: (1) To investigate the prevalence of common NCDs, such as hypertension, diabetes, dyslipidaemia, stroke, coronary heart disease (CHD), asthma, chronic obstructive pulmonary disease (COPD) and cancer; (2) To describe the dynamic changes in NCD prevalence and evaluate their long-term effects on individual health outcomes such as HRQoL across a long follow-up; (3) To evaluate the health effects of the anthropometric index, sociodemographic factors and living habits on NCDs; (4) To identify discriminative and predictive biomarkers that warrant further investigation for realising an early diagnosis, predicting risks and monitoring the recurrence of NCDs; and (5) To improve traditional preventive measures against NCDs in southern rural areas in Jiangsu.

COHORT DESCRIPTION

Cohort creation

Liyang, referred to as Liyang city (county-level city), is actually a county-level administrative unit in China's administrative division, with the urban and rural areas representing 20.28% and 79.72% of the population, respectively, based on the census results of Jiangsu Province in 2020 (online supplemental table 1). In 1986, The State Council of China decreed that 'counties' that met all of the following criteria could be called 'cities'. First, the total population reaches more than 500 000. In addition, the county government is located in towns with a non-agricultural population of more than 120 000. Second, the annual gross national product reaches more than ¥400 million. On 8 December 1990, Liyang City met all the above criteria and was officially called Liyang 'City'. In China, such county-level administrative units are often called 'county-level cities'. As an economically developed county-level city located at the junction of Jiangsu, Zhejiang and Anhui, and in the Yangtze River Delta, Liyang City was ranked among the top 10 counties and districts in Jiangsu Province, and its comprehensive strength, including economy, environment, and science and technology, ranked 25th among the top 100 counties in China in 2020. In the same year, the Gross Domestic Product (GDP) of Liyang had reached $17.73

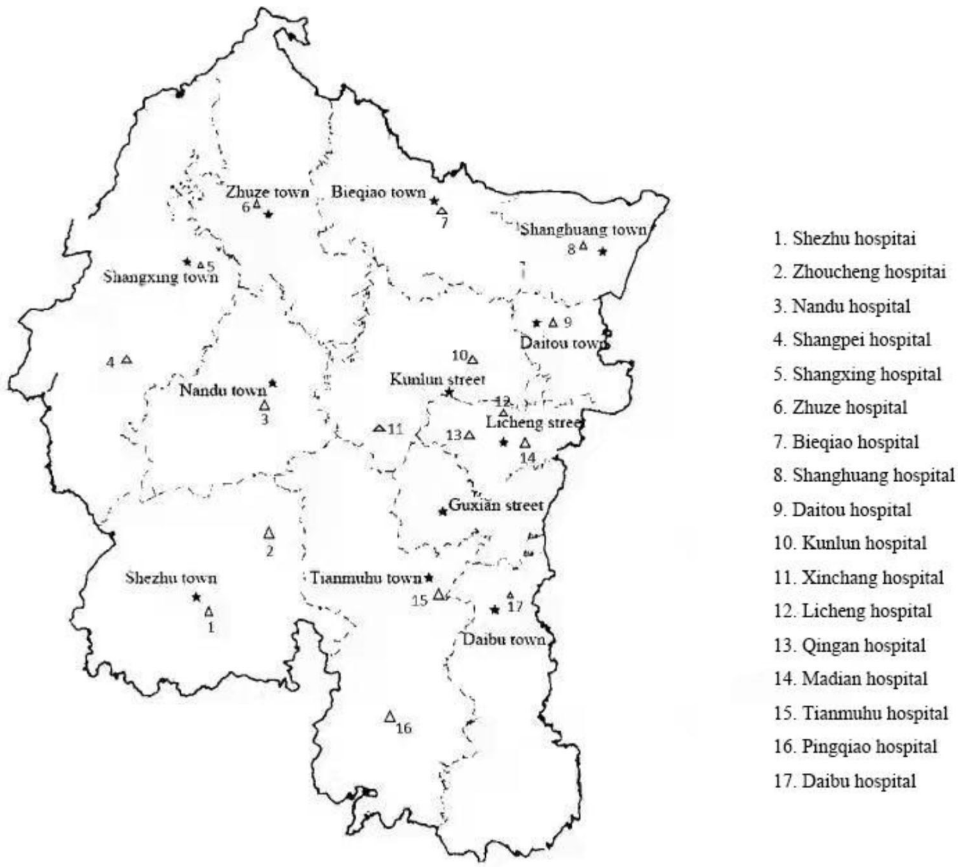

Figure 1 Locations of the 12 towns (streets) with 17 health centres in the Liyang Study. Note: The Licheng town has three health centers, Tianmuhu, Shangxing and Shezhu town have two heath centers,respectively, and the other 8 towns have only one health centers, respectively.

billion, which is a leading level in China. Also, residents' per capita disposable income had reached \$7090, far higher than the national average (\$5056). Despite this, the resource imbalance in Liyang still exists, mainly from a medical, education and income standpoint (online supplemental table 2). The per capita disposable income of urban residents (\$8784) is 1.84 times that of rural residents (\$4763). Besides, the GDP gap among the towns is also large, with the highest GDP 7.28 times the lowest.

To address the urgent need for understanding the prevalence and risk factors of NCDs in resource-imbalanced regions, the baseline survey of the Liyang Study was launched from March 2019 to June 2020 in Liyang city, Jiangsu Province. Samples were randomly selected from the general population using multistage stratified cluster sampling. Considering the consistency of residents, population stability and local medical conditions, all the towns and streets in Liyang city were selected as recruitment points, and the health centres (17 in total) responsible for daily healthcare in each town or street were selected as survey points (figure 1). Note: administrative divisions in China include provinces, cities, counties(districts) and towns(streets). Towns are the same level as streets and they are the most basic administrative units in China. Towns belong to rural areas, and residents living in them are mostly residents with rural household registrations, while streets belong to urban areas, and residents living in them have both urban and rural household registrations. The detailed population information of each town or street in Liyang city is shown in online supplemental table 1. We preliminarily planned to conduct a large-scale survey covering 10 000 people among the general population of Liyang city. First, to ensure a sufficient sample size and facilitate implementation and calculation, a sample of approximately 600 cases was planned for the survey at each survey site (health centre). However, due to the loss of address, refusing investigation of some urban residents, and some residents with rural household registrations living in the urban district, the sample size collected in several survey sites did not reach the planned 600 cases, so additional investigation on 261 alternative subjects was carried out at other survey sites to fill the gap. Specifically, because the sample size collected by several survey sites did not reach the requirements of the original plan, however, to ensure that the total number of people reached the expected target (a large cohort of more than 10 000 people), we randomly selected alternative participants (261 cases in total) from villages or communities belonging to the remaining survey sites (a total of 11 health centres recruited with a sample size of more than 600 shown in online supplemental table 1) whose participants are actively cooperating with this survey in

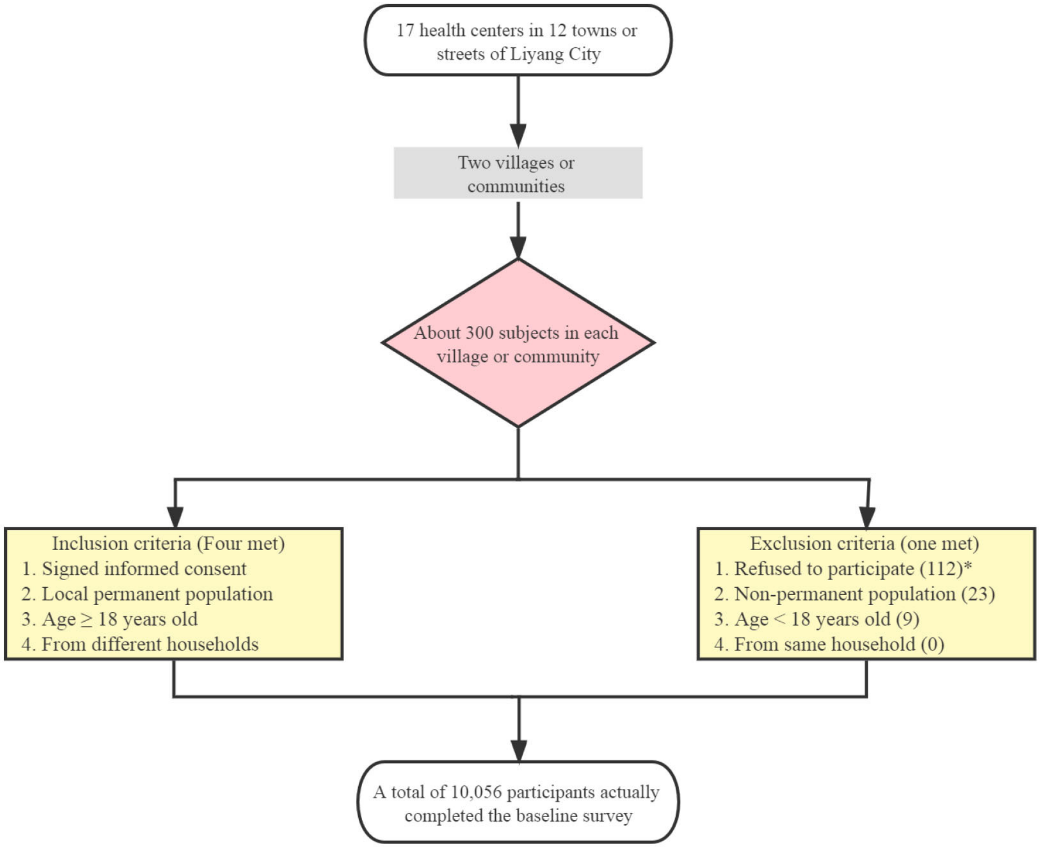

Figure 2 Flow diagram of baseline sampling in the Liyang Study. *Reasons for refusing to participate in the baseline survey included refusing to sign informed consent, failing a medical examination, and being physically unable to participate.

the quality control stage. Next, two administrative villages (belonging to the town) or communities (belonging to the street) in the vicinity of each health centre were randomly selected, and approximately 300 cases were planned to be randomly selected from each village or community as the respondents. Thus, 10 200 respondents were recruited for the survey before the questionnaire was issued, with an average sample size of 600 per survey site (ranging from 443 to 633) (online supplemental table 1). A total of 144 individuals were excluded from the study because they refused to sign informed consent (n=82), failed a medical examination (n=26), were physically unable to complete the study (n=4), were not permanent residents (n=23) or were younger than 18 years (n=9). Overall, 10 056 permanent residents (living for more than 6 months before the survey) from different households, aged ≥18 years, who had signed written informed consent, completed the baseline survey with a response rate of 98.6%. The sampling process and specific inclusion and exclusion criteria are illustrated in figure 2.

Data collection

The Liyang Study collected a large amount of information through questionnaire surveys, including potential confounding variables, mainly sociodemographic characteristics (such as age, sex, nationality, residence status, marital status, education level, living status, and family annual income and expenditure). The primary exposure variables were as follows: health-related behaviours (smoking, drinking, physical activity level, drinking tea and occupational exposure history), activities of daily living, instrumental activities of daily living to assess physical function status, abbreviated mental test to evaluate cognitive function, and Food Frequency Questionnaire (FFQ) to measure dietary habits. Moreover, the possible outcome variables included physician-diagnosed NCDs (ie, the symptoms persisted or related medical treatment continued over the past year), awareness of NCD-related knowledge, self-reported medical history and HRQoL as measured on the EuroQol 5-Dimensions 5-Levels validated by the Chinese population.[13] Detailed information on the questionnaire is presented in table 1.

Table 2 presents the details of the physical examinations and laboratory tests. The physical examination consisted of self-reported height, weight and waist circumference, and double measurements of height, weight, waist circumference and blood pressure using an international anthropometric instrument. While measuring height, the subjects removed their shoes and leaned against a calibrated wall upright. Waist circumference was measured at a level of 1.0 cm above the navel. An electronic weight scale (G&G TC-200K, USA) was used to measure subjects' weight according to the instructions. After an overnight fast, intravenous catheters were inserted into the antecubital vein for blood sampling by medical professionals.

Table 1 Summary of data collected for the Liyang Study

Component	Measurements
Sociodemographic data	Sex Ethnicity Residence status (urban/rural) Age (identity card number) Education level Marital status and history Employment status Living status (solitary or otherwise) Number of permanent household residents Family annual income
Health-related behaviour data	Smoking status Drinking status Drinking tea status Physical activities Diet status (salt, animal oil, vegetable oil and water) Occupational exposure history (type and duration)
Diseases data Physician-diagnosed NCDs 1. Prior medication history 2. Age of diagnosis 3. Family history 4. Awareness of NCD-related knowledge	Hypertension Diabetes Dyslipidaemia Stroke CHD COPD Asthma Cancer
Self-reported medical history (yes/no)	Chronic pulmonary diseases Hepatic diseases Heart diseases Kidney diseases Digestive diseases Mental health issues Memory-related diseases Arthritis or rheumatism
Scale data	
Measurement of dietary habits	FFQ Scale
Measurement of HRQoL	EQ-5D-5L Scale
Measurement of cognition function	AMT Scale backwards from 20 to 1, and recall of an address
Measurement of physical function (somatic aspects)	ADL Scale
Measurement of physical function (aspects of daily life)	IADL Scale

ADL, activities of daily living; AMT, abbreviated mental test; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; EQ-5D-5L, EuroQol 5-Dimensions 5-levels; FFQ, Food Frequency Questionnaire; HRQoL, health-related quality of life; IADL, instrumental activities of daily living; NCDs, non-communicable diseases.

Blood samples were drawn into a 5 mL serum separator tube (non-anticoagulant blood, yellow cap) and a 5 mL EDTA-K2 anticoagulation tube (purple cap). Blood samples were kept at room temperature for 1 hour before centrifugation at 3000 rpm for 15 min (TM48). The serum supernatant was used to measure blood glucose, total cholesterol, triglycerides, high-density lipoprotein and low-density lipoprotein (MINDRAY BS-820). After centrifugation, the anticoagulated whole blood was separated into plasma, white blood cells and red blood cells. The plasma and white blood cells were stored at −80°C and forwarded to measure gene expression. In addition, the white blood cells were centrifuged at 2000 rpm for 2 min before storage. All separated blood samples were sent to the Center for Disease Control and Prevention (CDC) within 24 hours and stored at −80°C (Haier DW-86L338J). The details of the flow chart for blood separation and detection of biochemical indices are shown in figure 3.

Data collection methods

Trained investigators conducted the baseline face-to-face interview. The baseline data in the Liyang Study were collected in two stages. In the first stage, we issued informed consent forms for the investigation in advance and set up centralised survey sites in local health centres. Subsequently, we required the respondents to carry their unique identification (ID) cards to the corresponding survey sites for inperson questionnaire surveys, which were conducted by trained and qualified interviewers. Professional physicians were responsible for physical measurements and blood sample collection. In the second stage, if residents could not participate in the centralised investigation, we arranged for interviewers to visit the participants' households for supplementary investigations, after obtaining informed consent from the respondents. Additionally, the results of physical measurements and laboratory tests were reported to the respondents within 2 weeks of the survey.

Quality control

At the beginning of the investigation, strict quality control procedures ensured that the quality of the study was in accordance with the unified standard. Prior to the baseline interview, a quality control committee was formed, comprising staff from Soochow University and Liyang CDC, and a quality control manual was developed by the committee. According to the manual, specific quality control measures were divided into two sections: primary and secondary. The primary quality control implemented by Soochow University involved checking the completeness and accuracy of the questionnaire and establishing the quality control level of the laboratory. Secondary quality control implemented by the Liyang CDC was to recheck all questionnaires in each town (street). For any errors detected, such as missing items, logical errors or incomplete filling, the respondents will be surveyed again: the investigator will communicate with the respondents by phone or door to door to ask them to fill the problematic questionnaire again. And then we will revise the wrong items according to the results from the repeat survey. In addition, a pilot study was conducted in three health centres (Daibu, Tianmuhu and Pingqiao) in December 2017 to assess aspects such as whether the questionnaire was easy to understand and if each question

Table 2 Summary of physical examination and laboratory measures collected at baseline in the Liyang Study

Variables	Number of measurements	Equipment used
Height	Twice	Height sitting height metre, TZG, China
Weight	Twice	Electronic weight scale, G&G TC-200K, USA
Waist circumference	Twice	Standard tape measure
Blood pressure	Thrice	Electronic sphygmomanometer, OMRON HBP1300, Japan
Fasting blood glucose	Once	Automatic blood biochemistry analyser, MINDRAY BS-820, China
Total cholesterol	Once	Automatic blood biochemistry analyser, MINDRAY BS-820, China
Triglycerides	Once	Automatic blood biochemistry analyser, MINDRAY BS-820, China
High-density lipoprotein	Once	Automatic blood biochemistry analyser, MINDRAY BS-820, China
Low-density lipoprotein	Once	Automatic blood biochemistry analyser, MINDRAY BS-820, China

was clear. After the pilot survey, we modified the language and items of the questionnaire according to feedback from the investigators.

Cohort follow-up

A follow-up survey will be conducted every 3 years, beginning in 2022. Like the baseline survey, all surviving study participants will be invited for repeated interviews, including questionnaire interviews and physical examinations. Although the Liyang Study collected blood samples from all participants at baseline, further blood samples, such as the genetics of patients with stroke, will be collected during follow-up only for specific substudies. Follow-up will be conducted both actively and passively. For active follow-up, a structured questionnaire using a computer-assisted personal interview system will be used to collect data. In addition, we will communicate with respondents not living (probably a small number) anymore at their original residence during the follow-up period, by phone in a timely manner. Although differences in data collection methods between baseline and follow-up may be a source of bias, a series of quality control measures implemented in the baseline stage effectively ensured the quality of data collected from paper questionnaires. First, investigators received uniform interviewing training before collecting data. Second, after data collection, we implemented strict quality control measures, including re-investigating missing, incomplete and logical errors items. Finally, at the data entry stage, we made a double entry and re-investigated the inconsistencies by phone. In addition to the steps involved in active follow-up, passive follow-up includes connecting to the primary Health Information System of local hospitals to collect basic healthcare information. Physician-diagnosed NCDs, medication history and family history of hypertension, diabetes, stroke, CHD, dyslipidaemia, COPD, asthma and cancer will be further identified by linking to the Chronic Disease Network Direct Report Management System of Changzhou city. All deaths during the follow-up period

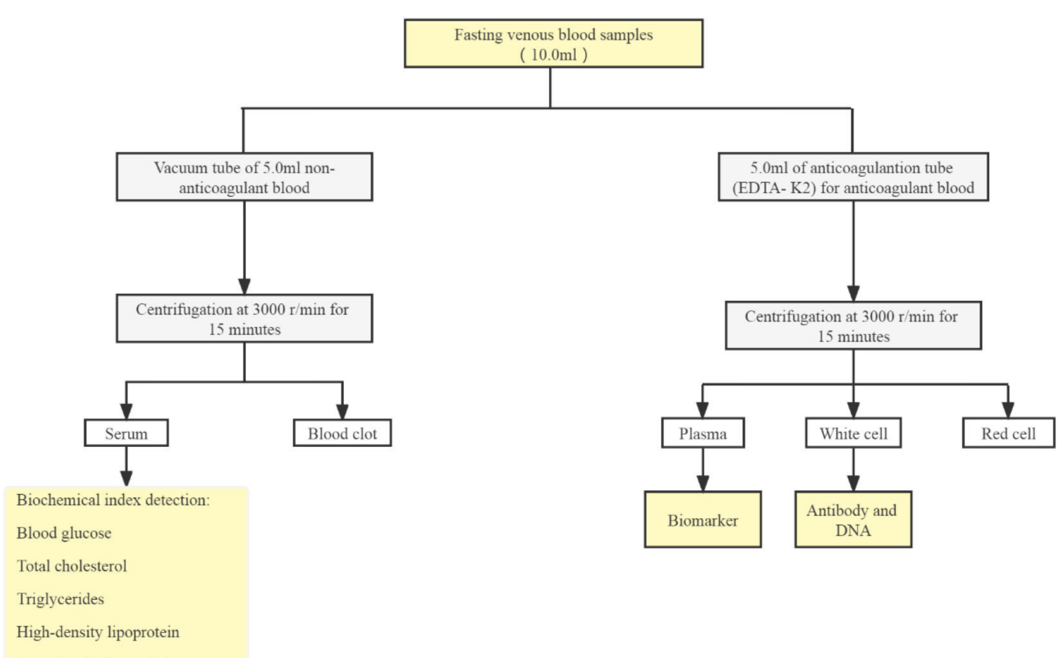

Figure 3 Flow chart of blood separation and detection of biochemical indexes.

will be checked according to cause-of-death classification through linkage with data from the Local Cause of Death Registration and Reporting System. Combined with passive follow-up, these existing systems will allow us to accurately gather basic sociodemographic information and track the incidence and prognosis of disease and death during follow-up. In addition, the under-representation of urban residents in this study may lead to selection bias. Hence, two measures should be taken into account to address this problem when analysing follow-up data. First, statistical methods such as standardised weighting or inverse probability weighting should be used to adjust the combined sample data according to the sampling ratio to lower selection bias.[14 15] Second, the rural and urban populations will be analysed separately so that the research findings can better reflect the different health status of the general population in Liyang city.

Findings to date

The baseline characteristics of the study participants were summarised as numbers with percentages and means±SD for categorical and continuous variables, respectively. Pearson's χ^2 analysis for categorical variables and Wilcoxon's rank-sum test for continuous variables were used to determine differences in baseline characteristics and the prevalence of NCDs according to sex. All data analyses were performed using SAS V.9.4 (SAS Institute, Cary, North Carolina, USA). A value of $p<0.05$ was considered statistically significant using two-sided tests.

Of the 10 056 individuals who participated in the Liyang Study, 47.08% (n=4734) were men, with a mean age of 49.9 years, and the vast majority (92.26%, n=9278) lived in rural areas. Approximately 31.51% (n=3172) had obtained high school or higher education. Most participants were married and employed (84.19%, n=8464 and 77.19%, n=7762, respectively). The proportion of participants with less than ¥50 000 per capita family annual income was 26.73% (n=2688). The prevalence of smoking and drinking was significantly higher in men than in women (50.91% for men, n=2410 vs 0.79%, n=42, for women and 41.38%, n=1959 for men vs 2.61%, n=139 for women, respectively). Overall, 52.83% (n=5313) of all participants had a normal BMI; the proportion of overweight and obesity was significantly higher in men than in women. Mean height, weight, waist circumference, systolic blood pressure, diastolic blood pressure, fasting blood glucose and triglycerides were significantly higher in men than in women (all $p<0.05$) (table 3).

The prevalence of the most frequent NCDs, based on self-reported and doctor-diagnosed data, in the Liyang Study is shown in table 4. The top three overall most prevalent chronic diseases in the study were hypertension (18.06%, n=1815), digestive diseases (7.88%, n=791), and arthritis or rheumatism (5.28%, n=530). Women had a significantly higher prevalence of diabetes and arthritis than men, while the opposite was true for chronic lung diseases such as COPD and heart diseases (all $p<0.05$).

Patient and public involvement

No participants were involved in the design and implementation of the study. In addition, all participants or their relatives were informed of the use of data in this study and the results of all examinations and tests were simply reported to each participant.

Strengths and limitations

The Liyang Study has several strengths. First, the most significant advantage of this study is that it provides referential data, based on a large sample size (n=10 056) and a wide range of variables, providing significant evidence for identifying and measuring the prevalence and risk factors of common NCDs. Another key advantage is that the field investigators and research team have significant fieldwork experience; moreover, professional medical institutions participated in the project, aided by the Liyang CDC and Soochow University, to provide project review and verification, as well as technical guidance. This multicollaborative team ensured the quality of our research. Second, the collection of a large number of biological samples at baseline, including whole blood, plasma, serum and white blood cells, will help effectively capture and determine relevant information about the genetics and variation of NCDs. The finding will have implications for disease prediction, prevention and awareness in China and in other countries. Finally, disease diagnosis and follow-up will be performed by linking various health record systems to realise real-time confirmation of patients' health status and medical history tracking.

It is also essential to highlight the limitations of the study. First, data collection on NCDs at baseline, especially mental and memory-related diseases, was self-reported, which may result in recall bias. However, effective researcher training, good field implementation, scientific and technical guidance, and good linkages with local health record systems will ensure the accuracy and reliability of the information. Moreover, studies have proven that self-reported information on NCDs possesses relatively good reliability.[16] Second, the study population was exclusively from Liyang city, Jiangsu Province, which may affect the generalisability of our results. However, the results of this large cohort study can, to some extent, indicate the main risk factors and prevalence of common NCDs in most developed areas of China. Third, due to the loss of address, refusal of investigation by some urban residents, and some residents with rural household registrations living in urban districts, the proportion of urban residents who completed the baseline survey was lower than that of urban residents in Liyang city, which may lead to bias in estimating the association of NCDs among the general population in Liyang city. Nevertheless, when we compare the age distribution and sex ratio of different age groups in the Liyang Study with those of the Liyang census (2020), we found that our cohort is quite similar to the entire population of Liyang city, supporting the representativeness of our study (online supplemental

Table 3 Baseline characteristics of the participants in the Liyang Study

Variables	Total (n=10 056)	Male (n=4734)	Female (n=5322)	T/χ^2	P value
Sociodemographic data					
Age groups, years	49.9±16.2	49.9±16.2	49.9±16.2	–0.190	0.851
18–30	1407 (13.99)	689 (14.55)	718 (13.49)	6.987	0.222
31–40	1699 (16.90)	769 (16.24)	930 (17.47)		
41–50	2408 (23.95)	1117 (23.60)	1291 (24.26)		
51–60	1905 (18.94)	928 (19.60)	977 (18.36)		
61–70	1411 (14.03)	656 (13.86)	755 (14.19)		
70+	1226 (12.19)	575 (12.15)	651 (12.23)		
Residence				0.185	0.667
Urban	778 (7.74)	372 (7.86)	406 (7.63)		
Rural	9278 (92.26)	4362 (92.14)	4916 (92.37)		
Education level				157.461	<0.001
Primary schools and below	3087 (30.70)	1189 (25.12)	1898 (35.66)		
Junior middle school	3797 (37.76)	1862 (39.33)	1935 (36.36)		
High school or similar	1904 (18.93)	1064 (22.48)	840 (15.78)		
Junior College and above	1268 (12.58)	619 (13.08)	649 (12.19)		
Annual household income (¥)				11.123	0.011
<50 000	2688 (26.73)	1228 (25.94)	1460 (27.43)		
50 000–99999	3316 (32.98)	1521 (32.13)	1795 (33.73)		
1 00 000–1 49 999	2231 (22.19)	1076 (22.73)	1155 (21.70)		
≥1 50 000	1821 (18.11)	909 (19.20)	912 (17.14)		
Employment status				843.243	<0.001
Paid employment	7762 (77.19)	4182 (88.34)	3580 (67.27)		
Student	176 (1.75)	76 (1.61)	100 (1.88)		
Unemployed	221 (2.20)	106 (2.24)	115 (2.16)		
Homemaker	1665 (16.56)	247 (5.22)	1418 (26.64)		
Retired	232 (2.31)	123 (2.60)	109 (2.05)		
Marital status				193.773	<0.001
Unmarried	954 (9.49)	580 (12.26)	374 (7.03)		
Married	8464 (84.19)	3925 (82.95)	4539 (85.30)		
Divorce	147 (1.46)	102 (2.16)	45 (0.85)		
Widow	483 (4.80)	123 (2.60)	360 (6.77)		
Others	5 (0.05)	2 (0.04)	3 (0.06)		
Health-related data					
Regular physical activities				0.094	0.759
Yes	6128 (60.94)	2892 (61.10)	3236 (60.80)		
No	3928 (39.06)	1842 (38.90)	2086 (39.20)		
Smoking statusSmoking status				3915.137	<0.001
Current	2452 (24.38)	2410 (50.91)	42 (0.79)		
Former	291 (2.89)	276 (5.83)	15 (0.28)		
Never	7313 (72.72)	2048 (43.26)	5265 (98.93)		
Drinking status				2280.887	<0.001
Current	2098 (20.86)	1959 (41.38)	139 (2.61)		
Never	7958 (79.14)	2775 (58.62)	51.83 (97.39)		
Physical examination data					

Continued

Table 3 Continued

Variables	Total (n=10 056)	Male (n=4734)	Female (n=5322)	T/χ^2	P value
BMI (kg/m^2)	23.97±7.48	24.34±9.76	23.65±4.56	4.620	<0.001
<18.5	328 (3.26)	120 (2.53)	208 (3.91)	60.763	<0.001
18.5–23.9	5313 (52.83)	2361 (49.87)	2952 (55.47)		
24.0 to 27.9	3440 (34.21)	1782 (37.64)	1658 (31.15)		
≥28.0	975 (9.70)	471 (9.95)	504 (9.47)		
Height (cm)	163.7±8.5	169.4±7.1	158.6±6.1	82.370	<0.001
Weight (kg)	64.2±12.8	69.5±14.2	59.4±9.2	42.710	<0.001
Waist circumference (cm)	82.2±8.8	84.7±8.4	80.0±8.6	27.810	<0.001
Systolic blood pressure (mm Hg)	128.0±15.7	129.4±14.5	126.7±16.6	8.460	<0.001
Diastolic blood pressure (mm Hg)	80.7±9.9	81.8±9.1	79.6±10.5	11.350	<0.001
Laboratory examination data					
Fasting blood glucose (mmol/L)	5.4±1.6	5.5±1.6	5.4±1.6	2.020	0.043
Total cholesterol (mmol/L)	5.0±1.2	4.9±1.3	5.0±1.0	–2.310	0.021
Triglycerides(mmol/L)	1.8±1.5	1.9±1.7	1.7±1.3	5.760	<0.001
High-density lipoprotein (mmol/L)	1.5±0.4	1.4±0.5	1.5±0.4	–8.330	<0.001
Low density lipoprotein (mmol/L)	2.9±0.8	2.9±0.8	2.9±0.8	–0.050	0.963

Note: Continuous data are presented as mean and SD (mean±SD), and categorical data are presented as number and percentage (n (%)).
BMI, body mass index.

table 3). Furthermore, we will account for rural and urban imbalances in analyses examining differences in NCD outcomes. Finally, in a cohort with an average age of 50 years, early life data for most participants will not be available. There is growing evidence that early life exposure may have long-term effects on risk of chronic diseases in adulthood.[4 17] However, the high prevalence and incidence of common NCDs among middle-aged

Table 4 Baseline prevalence of most frequent chronic diseases by sex group in the Liyang Study

Variables	Total (n=10 056)	Male (n=4734)	Female (n=5322)	χ^2	P value
Hypertension	1815 (18.06)	887 (18.75)	928 (17.45)	2.867	0.090
Diabetes	499 (4.97)	209 (4.42)	290 (5.46)	5.768	**0.016**
Dyslipidaemia	325 (3.23)	148 (3.13)	177 (3.33)	0.321	0.571
Stroke	185 (1.84)	91 (1.92)	94 (1.77)	0.338	0.561
CHD	173 (1.72)	80 (1.69)	93 (1.75)	0.049	0.825
COPD	114 (1.13)	65 (1.37)	49 (0.92)	4.584	**0.032**
Asthma	77 (0.77)	41 (0.87)	36 (0.68)	1.186	0.276
Cancer	110 (1.09)	49 (1.04)	61 (1.15)	0.286	0.593
Chronic pulmonary diseases	231 (2.30)	135 (2.86)	96 (1.81)	12.235	**<0.001**
Hepatic diseases	63 (0.63)	38 (0.80)	25 (0.47)	4.458	**0.035**
Heart diseases	222 (2.21)	104 (2.20)	118 (2.22)	0.005	0.942
Kidney diseases	66 (0.66)	31 (0.66)	35 (0.66)	0.001	0.985
Digestive diseases	791 (7.88)	349 (7.38)	442 (8.32)	3.031	0.082
Mental health issues*	116 (1.16)	53 (1.12)	63 (1.19)	0.092	0.762
Memory-related diseases†	92 (0.92)	43 (0.91)	49 (0.92)	0.005	0.947
Arthritis or rheumatism	530 (5.28)	209 (4.42)	321 (6.04)	13.153	**<0.001**

Statistical significance is shown in bold. Categorical data are presented as number and percentage (n (%)).
*Mental health issues mainly refers to schizophrenia, paranoid mental disorder.
†Memory-related diseases mainly include dementia (eg, Alzheimer's disease, Parkinson's disease).
CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease.

and elderly people will enable our study to provide sufficient information for future analysis.

Collaboration

All data collected are stored at the Soochow University, Soochow Medical College, School of Public Health. The Liyang Study encourages collaboration to maximise the use of data and sample information. Although the research database is currently not freely available in the public domain, as it contains sensitive information, any ideas and suggestions on potential collaborations are welcome. Researchers interested in collaborating can contact the authors via shenyueping@suda.edu.cn.

Acknowledgements The authors thank all study participants and staff of Soochow University, Shangxing Community Hospital, Zhuze Community Hospital, Bieqiao Community Hospital, Shanghuang Community Hospital, Nandu Community Hospital, Kunlun Community Hospital, Daitou Community Hospital, Licheng Community Hospital, Guxian Community Hospital, Shezhu Community Hospital, Tianmuhu Community Hospital, Daibu Community Hospital, and Liyang Center for Disease Control and Prevention.

Contributors Study design: WH, LZ, DH, SX, YQ, SL, CK and YS. Data collection and management: WH, NS, QH, LZ, XC, JC, TL, ZF, QH, LP, LC, YW, YS. Data analyses: WH, SL. Guarantor: YS. All authors were involved in manuscript preparation, and all authors read and approved the final manuscript. All authors meet the ICMJE criteria for authorship. All authors have participated actively in this study and agree to the content of the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (project number 81973143), the Liyang Chronic Disease Risk Factor Monitoring Cohort Study (project number P113911618) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Ethics Review Committee of Soochow University (reference number: SUDA20211025H02). All research procedures were approved by the Life Sciences Ethics Committee of Soochow University.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data availability statement:The research database contains sensitive information and therefore is not yet freely available in the public domain. However, upon reasonable request researchers may contact the corresponding author via email (shenyueping@suda.edu.cn) for cooperation.

ORCID iDs
Wei Hu http://orcid.org/0000-0003-3526-5773
Chaofu Ke http://orcid.org/0000-0003-3188-4307

REFERENCES

1 Yang G, Wang Y, Zeng Y, *et al*. Rapid health transition in China, 1990-2010: findings from the global burden of disease study 2010. *Lancet* 2013;381:1987–2015.
2 Zhou M, Wang H, Zhu J, *et al*. Cause-specific mortality for 240 causes in China during 1990-2013: a systematic subnational analysis for the global burden of disease study 2013. *Lancet* 2016;387:251–72.
3 Tan Z, Liang Y, Liu S, *et al*. Health-related quality of life as measured with EQ-5D among populations with and without specific chronic conditions: a population-based survey in Shaanxi Province, China. *PLoS One* 2013;8:e65958.
4 Liu X, Mao Z, Li Y, *et al*. Cohort profile: the Henan rural cohort: a prospective study of chronic non-communicable diseases. *Int J Epidemiol* 2019;48:1756–1756j.
5 Chen W-W, Gao R-L, Liu L-S, *et al*. China cardiovascular diseases report 2015: a summary. *J Geriatr Cardiol* 2017;14:1–10.
6 Vandenberghe D, Albrecht J. The financial burden of non-communicable diseases in the European Union: a systematic review. *Eur J Public Health* 2020;30:833–9.
7 Li X, Lu J, Hu S, *et al*. The primary health-care system in China. *Lancet* 2017;390:2584–94.
8 Niu SR, Yang GH, Chen ZM, *et al*. Emerging tobacco hazards in China: 2. early mortality results from a prospective study. *BMJ* 1998;317:1423–4.
9 Chen Z, Lee L, Chen J, *et al*. Cohort profile: the Kadoorie study of chronic disease in China (KSCDC). *Int J Epidemiol* 2005;34:1243–9.
10 Zhou T, Guan H, Yao J, *et al*. The quality of life in Chinese population with chronic non-communicable diseases according to EQ-5D-3L: a systematic review. *Qual Life Res* 2018;27:2799–814.
11 Wang A, Zhang J, Li J, *et al*. Relationship between time-dependent proteinuria and risk of stroke in population with different glucose tolerance status. *J Am Heart Assoc* 2020;9:e015776.
12 Zhao Y, Hu Y, Smith JP, *et al*. Cohort profile: the China health and retirement longitudinal study (CHARLS). *Int J Epidemiol* 2014;43:61–8.
13 Luo N, Liu G, Li M, *et al*. Estimating an EQ-5D-5L value set for China. *Value Health* 2017;20:662–9.
14 Haneuse S, Schildcrout J, Crane P, *et al*. Adjustment for selection bias in observational studies with application to the analysis of autopsy data. *Neuroepidemiology* 2009;32:229–39.
15 Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. *Acta Obstet Gynecol Scand* 2018;97:407–16.
16 Midthjell K, Holmen J, Bjørndal A, *et al*. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trøndelag diabetes study. *J Epidemiol Community Health* 1992;46:537–42.
17 Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 2002;31:285–93.