International Journal of Epidemiology, 2015, 810–818
doi: 10.1093/ije/dyv013
Advance Access Publication Date: 2 March 2015
Cohort Profile

Cohort Profile

Cohort Profile: The Shanghai Men's Health Study

Xiao-Ou Shu,[1]* Honglan Li,[2] Gong Yang,[1] Jing Gao,[2] Hui Cai,[1] Yumie Takata,[1] Wei Zheng[1] and Yong-Bing Xiang[2]

[1]Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, TN, USA and [2]Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

*Corresponding author. Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, 2525 West End Avenue, Suite 600, Nashville, TN 37203-1738, USA. E-mail: xiao-ou.shu@vanderbilt.edu

Accepted 22 January 2015

Abstract

The Shanghai Men's Health Study (SMHS) is a population-based cohort study of 61 480 men aged 40–74 years, launched in 2002 in urban Shanghai to investigate the contribution of lifestyle/environmental factors and genetic susceptibility to cancer and other non-communicable diseases (NCDs). At baseline, trained interviewers collected detailed information on personal and dietary habits, occupational/medical history and physical activity, and took anthropometric measurements (response rate: 74%). Blood, urine and DNA were collected from 75%, 89% and 89% of participants, respectively. The cohort has been followed up through a combination of in-person surveys every 3–4 years and annual record linkage with cancer and vital statistics registries. Response rates for in-person follow-up surveys were over 91% and coverage for mortality nearly 100%. SMHS participants have a high smoking rate (58.6%) and moderate alcohol-drinking rate (29.3%), but low obesity rate (2.6%). They have a low calorie intake from fat (16.2% of total calorie intake) and protein (16.4%), high calorie intake from carbohydrates (67.4%), and high intake of soy food, cruciferous vegetables and fish (156.5, 110.6 and 51.7 g/day, respectively). With its unique exposure pattern and wealth of data and biological samples, the SMHS is well positioned for long-term research into NCD aetiology and prognosis. Information about accessing the SMHS resources can be found at: http://www.mc.vanderbilt.edu/swhs-smhs/.

Key Messages

- Chinese men living in Shanghai have a unique NCD spectrum and pattern of exposures which facilitate epidemiological investigation of the aetiology of these diseases that are difficult to study in cohorts conducted in Western populations.
- The Shanghai Men's Health Study (SMHS), a population-based cohort study, has high response rates at study enrolment and in the follow-up surveys. Extensive, high-quality survey and record linkage data have been collected. A blood or buccal cell sample and a urine sample were collected for nearly 90% of study participants, providing a unique opportunity for NCD research.
- The SMHS has shown that lifestyle factors—including diet, physical activity and supplement use—play an important role in NCD development and mortality, and has contributed to the identification of multiple genetic susceptibility biomarkers for cancer, body mass index, and diabetes.

Why was the cohort set up?

Each year, 36 million people die of non-communicable diseases (NCDs) globally, 80% (29 million) of them in developing countries.[1] The health burden of NCDs in developing countries is expected to continue rising,[2] mainly due to population ageing and changes in lifestyle and environmental exposures.

China, home to one-fifth of the world's population, has experienced rapid economic development and major environmental and lifestyle changes as well as an increase in the prevalence of many known risk factors for NCDs, over the past few decades. For example, the prevalence of overweight and obesity in China doubled among women and tripled among men between 1989 and 1997.[3] Type 2 diabetes (T2D) prevalence rose from 1.9% to 5.6% between 1993 and 2003.[4] Similarly, the prevalences of hypertension, dyslipidaemia, and diabetes among Chinese adults has increased substantially, approaching 26.1%, 53.6% and 5.2%, respectively, in 2000 and 2001 for individuals between the ages of 30 and 74.[5] In Shanghai, one of the largest cities in China, the age-adjusted incidence rates of colon cancer increased by 147% in men and 161% in women between 1973 and 1977 and between 1998 and 2000, respectively.[6] During the same period, dramatic increases in the incidence of cancers of the breast (69%), prostate (281%), uterus (140%), pancreas (79% in men and 90% in women) and brain (42% in men and 67% in women) have been witnessed.[6] In China, cardiovascular disease (CVD) and cancer have surpassed infectious diseases as the leading causes of death, and NCDs accounted for an estimated 82.9% of total deaths and 70% of disability-adjusted life-years lost in China in 2000.[7] As westernization continues and the large population of baby boomers begins to age, China will face an increased burden of NCDs. This transition in the epidemiology of NCDs in China presents an enormous challenge to both public health researchers and policy makers, and offers an exceptional research opportunity to investigate the contribution of environmental and lifestyle factors and their interaction with genetic susceptibility in NCD development.

Supported by the National Cancer Institute, we launched the Shanghai Men's Health Study (SMHS), a population-based cohort study, in 2002 for a comprehensive and long-term investigation of NCD aetiology and progression. The study protocols were approved by the Institutional Review Boards of Vanderbilt University and the Shanghai Cancer Institute. Informed consent was obtained from all participants.

Who is in the cohort?

Eight typical neighbourhood communities in the Changning district of urban Shanghai were chosen for the SMHS. These communities are similar to the general population of urban Shanghai in terms of demographic characteristics and cancer incidence rates. Recruitment for the SMHS was initiated in spring 2002. All men aged 40–74 who were permanent residents of these communities were approached for the study by trained interviewers. By June 2006, 83 033 potential eligible men had been approached for the study; 61 480 were enrolled, an overall response rate of 74.0%. The response rate increased steadily with age of study participants, from 65.9% for the 40–44 year-old age group to 68.6%, 71.6%, 76.2, 81.3%, 85.4% and 86.4% for each additional 5-year increment in age.

The baseline demographics of the study participants are presented in Table 1. The average age of participants was 54 at study enrolment. Approximately two-thirds (68.7%) of participants had received a middle- or high-school education, 23.3%, above high-school and 6.6%, less than middle-school education. Per capita monthly income was between 500 and 2000 Chinese yuans for most study participants (77.7%). The average number of jobs

Table 1. Baseline demographic characteristics of the Shanghai Men's Health Study participants

Characteristics	Total ($N = 61480$)	Remain in study ($N = 60789$)	Lost to follow-up ($N = 691$)	p-Value*
Average age (mean ± SD, years)	54.88 ± 9.74	54.89 ± 9.75	53.92 ± 9.13	0.009
Education				
No formal education	1.12%	1.13%	0.29%	
Elementary	5.53%	5.56%	2.89%	
Middle school	33.08%	33.13%	28.51%	
High school	35.56%	35.54%	37.48%	
Technical school/college or above	23.32%	23.24%	19.95%	
Unknown	1.41%	1.41%	0.87%	<0.001
Per capita monthly income (yuans)				
<500	12.59%	12.59%	12.41%	
500 ≤ income < 1000	42.59%	42.62%	39.56%	
1000 ≤ income < 2000	35.11%	35.13%	33.87%	
2000 ≤ income < 3000	7.49%	7.46%	10.66%	
≥3000	2.22%	2.20%	3.51%	0.009
Longest job				
Professional/technical/administrative	26.56%	26.54%	28.26%	
Clerical/commercial/services	21.94%	21.93%	22.90%	
Manufacture/manual labour	51.50%	51.53%	48.84%	0.365
Number of jobs ever held (mean + SD)	3.18 + 1.44	3.19 ± 1.44	3.01 ± 1.41	0.001
Years worked in the longest job (mean + SD)	20.70 ± 8.79	20.71 ± 8.79	20.30 ± 8.88	0.223
Family history of cancer	28.34%	28.34%	28.36%	0.987

*p-Value for comparison of subjects who remain in study with those lost to follow-up.

ever held by participants was 3.2. The longest duration of job held was 21 years on average; 51.5% of jobs were manufacturing or manual, 26.6% professional/technical/administrative and the remaining 21.9% clerical/commercial/service.

How is the cohort followed up?

The cohort has been followed up by a combination of regular record linkage with population-based registries and in-person surveys. To identify new cancer cases in the SMHS, the files of the population-based Shanghai Cancer Registry (SCR) in our study district (Changning District) are searched monthly. In addition, we conduct annual record linkages to the Shanghai municipal SCR to identify additional cancer cases (primarily for those relocated outside the study district after the baseline survey), the Shanghai Vital Statistics Registry to gather information on vital status and cause of death and the Shanghai Residential Registry to update current residential information. All possible matches are checked manually and verified through home visits. All cancer diagnoses are further verified by reviewing medical records from hospitals.

In addition, every 3 or 4 years we conduct in-person follow-up surveys to update and collect new exposure information and to collect information on the interim health history including cancer, cardiovascular disease, stroke, bone fracture and other chronic diseases that occurred since the last in-person contact. We have completed two in-person follow-up surveys in the SMHS with response rates of 97.6% and 91.9%, respectively. The third in-person follow-up survey, initiated in 2012, is ongoing. Participants who were lost to follow-up were younger, slightly less likely to have received a college education, and more likely to have a higher income than those who remained in the study (Table 1). However, subjects who were lost to follow-up did not differ from those remaining in the study for most of the lifestyle factors that are known to be related to NCD (Table 2).

What has been measured?

Exposure information

Presented in Table 3 is a summary of the information collected in the SMHS. At study enrolment, in-person

Table 2. Selected baseline characteristics of Shanghai Men's Health Study participants

Characteristics	Total ($N = 61480$)	Remaining ($N = 60789$)	Lost to follow-up (N = 691)	p-Value*
BMI (mean ± SD)	23.7 ± 3.1	23.7 ± 3.1	23.7 ± 3.1	0.84
Overweight (BMI ≥ 25–< 30 kg/m^2)	30.5%	30.5%	30.1%	0.81
Overweight or obese (BMI ≥ 25)	33.1%	33.1%	33.3%	0.91
Obese (BMI ≥ 30)	2.6%	2.6%	3.2%	0.28
WHR (mean ± SD)	0.90 ± 0.06	0.90 ± 0.06	0.90 ± 0.06	0.53
Waist circumference (cm)	85.1 ± 8.7	85.1 ± 8.7	85.1 ± 8.4	0.95
Currently smoking (%)	58.6%	58.6%	59.2%	0.77
Cigarettes/day (mean + SD) among smokers	16.5 ± 8.8	16.5 ± 8.8	16.6 ± 8.8	0.86
1-14 cigarettes/day	40.5%	40.4%	42.9%	
≥15 cigarettes/day	59.5%	59.6%	57.1%	0.27
≤20 cigarettes/day	88.5%	88.5%	87.4%	
>20 cigarettes/day	11.5%	11.5%	12.6%	0.44
pack-years of smoking	24.5 ± 16.2	24.6 ± 16.2	23.9 ± 16.0	0.35
Currently drinking (any amount) (%)	29.3%	29.3%	28.5%	0.65
Heavy drinkers (≥ 30 g/day)	12.8%	12.8%	10.2%	0.04
Ethanol intake (g/day) (mean ±SD)	11.8 ± 25.1	11.8 ± 25.1	10.3 ± 22.4	0.12
Regular exercise (%)	35.5%	35.5%	36.3%	0.66
Energy intake (kcal/day)	1909 ± 485	1909 ± 485	1937 ± 498	0.12
% of energy from fat	16.2 ± 5.2	16.2 ± 5.2	16.5 ± 5.5	0.09
% of energy from carbohydrate	67.4 ± 7.4	67.4 ± 7.4	66.8 ± 7.6	0.06
% of energy from protein	16.4 ± 2.8	16.4 ± 2.8	16.6 ± 2.7	0.07
Red meat (g/day) (mean ± SD)	63.3 ± 44.8	63.2 ± 44.8	65.1 ± 45.5	0.29
Poultry (g/day) (mean ± SD)	16.0 ± 19.2	16.0 ± 19.2	16.7 ± 20.3	0.38
Fish (g/day) (mean ± SD)	51.7 ± 46.4	51.7 ± 46.4	52.6 ± 42.8	0.64
Total vegetables (g/day) (mean ± SD)	343.8 ± 192.9	343.7 ± 192.8	353.8 ± 202.0	0.21
Cruciferous vegetables (g/day) (mean ± SD)	110.6 ± 74.8	110.5 ± 74.8	114.6 ± 76.8	0.15
Allium vegetables (g/day) (mean ± SD)	16.4 ± 17.4	16.4 ± 17.3	17.6 ± 21.3	0.06
Soy food (g/day) (mean ± SD)	156.5 ± 119.1	156.5 ± 119.0	159.9 ± 123.0	0.45
Isoflavone (mg/day)	35.7 ± 23.9	35.6 ± 23.9	37.0 ± 25.4	0.15
Total fruit (g/day) (mean ± SD)	151.6 ± 126.1	151.5 ± 126.0	161.3 ± 134.0	0.04
Calcium supplement	4.7%	4.7%	5.8%	0.18
Multivitamin supplement	7.5%	7.5%	9.6%	0.05
Vitamin D supplement	0.5%	0.5%	0.6%	0.58
Fish oil supplement	6.8%	6.8%	7.4%	0.48

*p-Value for comparison of subjects who remain in study with those lost to follow-up.

interviews were administered at each study participant's home using a structured questionnaire which included these sections: (i) demographic background; (ii) medical history; (iii) personal habits; (iv) dietary habits; (v) physical activity; (vi) family history of cancer; (vii) occupational history; and (viii) weight history and height during adolescence and early adult years. In the dietary section of the questionnaire, we obtained information on usual dietary intake over the preceding 12 months and during adolescence. The validity and reliability of the food frequency questionnaire (FFQ) and physical activity questionnaire for the SMHS have been evaluated and reported previously.[8–10] In addition, weight, standing and sitting height and circumferences of the waist and hips were measured. Pulse rate and two blood pressure measurements were taken by interviewers at the time of interview in accordance with standard protocols.

During the in-person follow-up surveys, structured questionnaires were used to update and collect new exposure information, including lifestyle factors and dietary intake. Blood pressure was taken by the interviewers at the time of the survey. Another comprehensive dietary survey was carried out at the first follow-up survey.

In addition to the annual record linkage with cancer registry and vital statistics databases mentioned above, the SMHS has recently initiated a new record linkage with the Changning District Health Information System which has collected electronic medical and prescription records, clinical laboratory test results and hospital discharge information from all district hospitals ($n = 6$) and community

Table 3. Measurement content of in-person surveys and response rate of the Shanghai Men's Health Study

Survey (year)	Response rate (%)	Exposure assessment*
Baseline recruitment (2002–06)	74.0	Demographics, FFQ, personal habits (smoking, alcohol and tea consumption), supplement use, weight and medical history, occupation history, family history of cancer, CVD and diabetes. In-home anthropometrics (weight, height, waist/hip circumference) and blood pressure measurements
First follow-up (2004–08)	97.6	FFQ (85.0% response rate), benign or malignant tumours, prostate atrophy, diabetes, myocardial infarction, stroke, bone fracture and other common chronic diseases, tea consumption, smoking, weight and supplement use. In-home blood pressure and pulse measurements
Second follow-up (2008–11)	91.9	Benign or malignant tumours, prostate atrophy, diabetes, myocardial infarction, stroke, bone fracture and other common chronic diseases, supplement use, sleep habits, weight, depression, oral hygiene, use of Chinese medicine. In-home blood pressure and pulse measurements
Third follow-up (2012–)	N/A	Benign or malignant tumours, prostate atrophy, diabetes, myocardial infarction, stroke, bone fracture and other common chronic diseases, supplement use, smoking and tea consumption, cancer screening, weight, family history of chronic disease, hair loss, physical function, memory and hearing loss and social interaction. In-home blood pressure and pulse measurements

N/A, not yet available.

*Information for all the exposures was based on self-report except the anthropometrics, blood pressure and pulse measurements which were taken by interviewers at the participants' homes at times of interview.

health centres ($n = 10$) in our study districts since 2007. This new record linkage will greatly facilitate the corroboration of self-reported medical events and medication use, and opens new avenues for the SMHS to carry out pharmacoepidemiological and outcome research.

Biospecimens

At study enrolment, we collected a 10-ml blood sample from each of 46 244 willing study participants (75.1%), a buccal cell sample from 8562 of those who did not donate a blood sample (56.0%) and a spot urine sample from 54 769 (88.9%). All samples were kept at 4°C, processed within 6 h and stored at −75°C. At biospecimen collection, information was collected on date and time of sample collection, time of last meal, intake of selected foods, smoking and use of medications over the past 24 h and during the preceding week.

For cancer patients who receive surgical treatment, we attempt to obtain eight 5-μm and two 15-μm unstained tumour tissue sections from the hospital. Tissue sections are coated with a thin layer of wax to prevent oxidation and are stored in a cold room at 4°C. Post-diagnosis urine and saliva samples are collected from cancer patients whenever possible.

What has been found? Key findings and publications

Distribution of lifestyle factors. The smoking rate was high among SMHS participants: 69.6% ever smoked and 58.6% currently smoked. Among the smokers, 11.5% smoked more than a pack of cigarettes per day. About one-third (29.3%) of participants drank alcoholic beverages regularly, with 12.8% being heavy drinkers ($\geq$ 30 g ethanol/day, equivalent to $\geq$ 2 drinks/day). Two-thirds (67.1%) of participants drank tea regularly and 35.5% exercised regularly. The prevalence of overweight/obese (BMI $\geq$ 25 kg/m^2) was 33.1% and of obesity (BMI $\geq$ 30 kg/m^2) 2.6%. The mean energy intake was 1909 kcal/day, with fat, protein and carbohydrates contributing 16.2%, 16.4%, and 67.4% of total calorie intake, respectively. The average daily intake of total vegetables, soy food, and cruciferous and allium vegetables was 343.8, 156.5, 110.6 and 16.4 g, respectively. Red meat, poultry, and fish intakes averaged 63.3, 16.0 and 51.7 g/day.

Prevalence of common chronic diseases. At baseline survey, 29.9%, 8.5%, 7.5%, 6.3% and 3.8% of SMHS participants reported a history of hypertension, prostatic hypertrophy, cholelithiasis, diabetes or stroke, respectively. Information on new occurrences of these conditions was collected during follow-up surveys. Information on any form of hepatitis or fatty liver was collected at the second follow-up survey and was reported by 9.8% and 16.7% of participants, respectively. Among those diagnosed with hepatitis, 28.8% had hepatitis B virus infection, and 57.3% hepatitis A virus infection.

Top incident cancer and cause of death. As of 31 December 2011, 3133 incident cancer cases and 3796 deaths were documented after a median follow-up of

Table 4. Summary of major findings from the Shanghai Men's Health Study

Exposures	Major findings
Diet and nutrition	Vitamin E intake from diet and supplements: inversely associated with liver cancer risk[11]
	Tea drinking: inversely associated with colorectal cancer risk[12]
	Fruit consumption: inversely associated with stomach cancer, lung cancer and CHD risk[13,47]
	Crucifers, green leafy vegetable consumption: inversely associated with lung cancer risk, CHD and total mortality[15,47]
	Peanut consumption: inversely associated with CHD and total mortality[48]
	Fish intake: inversely associated with total, ischaemic stroke and diabetes mortalities[14,16]
	Red meat intake: positively associated with total and CHD mortalities[49]
	Soy food consumption: positively associated with CHD risk and biomarkers[50]
	Visceral adiposity: positively associated with CHD risk[51]
	Height: positively associated with cancer mortality but inversely associated with CVD mortality[52]
Physical activity	Moderate-intensity exercise related to reduced total, CVD and cancer mortalities[17]
Biomarkers	*H. pylori*: six *H. Pylori* antibodies found to be associated with stomach cancer risk[18]
	Folate: positively associated with colorectal cancer risk among men with possible preneoplastic lesions[20]
	C-reactive protein: positively associated with colorectal cancer risk[19]
Genetic factors	8 novel BMI-related genetic loci identified[24,25]
	A novel shared stomach/oesophageal cancer genetic locus identified[23]
	17 novel colorectal cancer-related genetic loci identified[21,22]
	9 novel diabetes-related genetic loci identified[53,54]

6.5 years. The most common cancer was lung cancer, accounting for 19.6% of all cancer cases, followed by cancers of the stomach (12.5%), colon (9.5%), liver (9.1%), prostate (8.7%), rectum (6.4%), kidney (4.7%), pancreas (3.7%), bladder (3.2%) and oesophagus (2.6%). Lung cancer was the leading cause of death in the SMHS, accounting for 13.0% of all deaths. Intracerebral haemorrhage, acute myocardial infarction, stomach cancer, liver cancer, cerebral artery occlusion, diabetes, pancreatic cancer, colon cancer and other forms of chronic ischaemic heart disease ranked the second to tenth most common causes of death.

Highlights of major findings. Over the past 12 years, the SMHS resources have contributed to over 40 institutional and international consortium studies and nearly 100 publications. Selected major findings are summarized in Table 4. Among the most significant contributions are identification of dietary cancer-inhibitory factors, lifestyle risk/protective factors for other NCDs, biomarkers for cancer risk assessment and genetic susceptibility markers for cancer and other chronic diseases. For example, we found that vitamin E intake, both from dietary sources and from supplements, was associated with reduced risk of liver cancer, one of the most lethal cancers, in a dose-response manner. This association was stronger among individuals with a history of liver disease, providing a possible basis for dietary intervention to prevent liver cancer.[11] In another report from the SMHS, we found that regular green tea consumption was inversely associated with overall risk of colorectal cancer; each 2-g increment in intake of dry green tea leaves per day (approximately equivalent to the amount of tea in a tea bag) was associated with a 12% reduction in risk.[12] Our study on stomach cancer found that increased fruit intake was associated with a 50% decreased risk of distal gastric cancer,[13] an association more evident among ever smokers than never smokers. We found that fish intake was related to a reduced risk of diabetes.[14] High consumption of cruciferous vegetables or fish, or participation in low-intensity exercise, was associated with lower incidence of diabetes and/or cardiovascular disease mortality.[15–17]

Using pre-diagnosis blood collected at baseline, we assessed the association of 15 antibodies to *Helicobacter pylori* (*H. pylori*) proteins in a nested case-control study of gastric cancer and found that an increasing number of sero-positive results to six *H. pylori* proteins (Omp, HP0305, HyuA, HpaA, CagA and VacA) may predict the risk of gastric cancer better than CagA alone, a known biomarker for gastric cancer. Compared with individuals with more than three sero-positive results to the six virulent proteins identified in this population, individuals with four to five sero-positive results were at a 2-fold increased risk, and individuals with sero-positive results to all six proteins had a 3.3-fold increase in risk for gastric cancer.[18] We found that high levels of C-reactive protein or folate in pre-diagnostic plasma samples were associated with an increased risk of colorectal cancer, and that these associations were more evident for subjects whose cancer was diagnosed closer to their blood draw.[19,20]

SMHS resources have also contributed to several genome-wide association studies which have identified novel genetic markers for colorectal, gastric and pancreatic

cancer as well as body mass index.[21–25] The SMHS is one of the major contributors to the Asian genome-wide association study (GWAS) consortium for colorectal cancer, that has identified 17 novel loci.[21,22] In addition, the SMHS has contributed to several consortium GWAS or pooling projects on significant NCD risk, mortality and risk factors, such as obesity, diabetes and hypertension.[23–31]

What are the main strengths and weaknesses?

The SMHS is unique in many ways, providing opportunities to investigate important aetiological hypotheses that cannot be investigated adequately in many other existing cohort studies.

Unique exposure patterns. Compared with participants of most cohort studies conducted in the USA, SMHS participants have an overall lower prevalence of obesity, lower percentage of energy intake from fat, higher consumption of carbohydrates, fish, soy foods and cruciferous and allium vegetables. However, the smoking rate is much higher among SMHS participants than in their US counterparts. Most of these factors have been shown or suggested to play an important role in cancer, CVD and diabetes, but findings are not entirely consistent.[32–35] The unique exposure patterns, repeated dietary surveys and biospecimens make the SMHS an invaluable resource for testing many aetiological hypotheses, especially those related to dietary protective factors that may be difficult to investigate adequately in other cohort studies.

Unique disease spectrums. Cancer incidence varies greatly worldwide.[1,36] Whereas prostate, lung and colorectal cancers are the three most common malignancies among US men, comprising > 58.7% of all cancer diagnoses in 2007,[37] they are much less common in Asian countries including China.[1] In Shanghai, the age-adjusted incidence rates of prostate and colorectal cancers in 2007 were 9.31/100 000 and 25.8/100 000, respectively, roughly 5.8% and 47.9% of the rates observed in the USA.[1,6,38] However, the incidence rates of these two cancers have increased rapidly among Chinese immigrants to the USA[30,39] as well as in Shanghai in the past three decades (a 148% and 281% increase in men from 1973–77 to 1998–2000). The smoking rate among US men has been < 30% since 2000, compared with > 60% for the past two decades among men in Shanghai. The incidence rates for several smoking-related cancers, including lung, pancreatic and bladder cancer, however, are paradoxically much lower among Chinese men (45.1, 7.3, 7.7/100 000, respectively) than among US men (82.7, 37.0, 13.6/100 000, respectively).[37,38] Although some of the differences could be attributed to a short duration of the epidemic of cigarette smoking in China, these striking contrasts also provide exceptional opportunities for investigating potential interactions between smoking and other lifestyle factors, as well as genetic factors, in the aetiology of these cancers. Similarly, gastric and liver cancer are among the most common malignancies globally, although relatively uncommon in the USA and other developed countries.[1] The SMHS provides a unique opportunity to investigate risk factors for these important cancers prospectively.

Other common NCDs. In the SMHS, diabetes, hypertension, hyperlipidaemia, stroke, fatty liver and cholelithiasis, a set of metabolic-related conditions that are typically tied to obesity, are common. However, the prevalence of obesity in the SMHS is only 2.6%. It has been suggested that Asians may be more susceptible to insulin resistance, given the same level of body fat, when compared with their Western counterparts.[31] The availability of standardly-measured anthropometrics and biological samples in the SMHS will allow a comprehensive investigation of the biological mechanisms underlying these metabolic diseases.

Biological samples. The SMHS has collected and stored blood and urine samples from most study participants, providing valuable resources for molecular epidemiological studies for NCD risk assessment and early detection. Notably, the SMHS is one of few existing cohort studies to have collected and stored urine samples for molecular epidemiological studies of urinary biomarkers. The value of urine samples has been increasingly recognized, since some biomarkers can only be reliably measured in urine. Furthermore, compared with plasma biomarkers, urine biomarkers often reflect relatively longer-term levels of exposure (e.g. phytoestrogens). Over the years, we have investigated multiple urinary biomarkers in relation to cancer risk,[40–46] and found a strong dose-response association of colorectal cancer with urinary prostaglandin E metabolites (PGE-M), the best biomarker of endogenous prostaglandin E_2 (PGE2) production.[42] The value of urinary biomarkers in NCDs will only increase as advances in molecular technology, such as metabolomics, become widely applied.

Limitations. Like most similar prospective cohort studies, the SMHS only collected biospecimens from its participants at one time point, which limits its ability to investigate changes in biomarkers of exposures or disease. This limitation will be ameliorated by our ongoing effort to collect second blood and urine samples from about 5000 SMHS participants. Except for cancer, information on

occurrence of other chronic diseases in the SMHS has been primarily based on unverified self-report. The newly-initiated record linkage with the Changning Health Information System will allow for verification of self-reported information on major chronic diseases, and offer the opportunity to conduct pharmacoepidemiological investigations.

In summary. The SMHS has collected extensive, high-quality survey data and biological samples, providing excellent opportunities to prospectively test a wide range of hypotheses, and to investigate gene-environment interactions and discover biomarkers for NCD risk assessment and early detection. The SMHS employs rigorous methodology, implements repeated exposure assessments and is one of very few cohort studies to collect tumour tissue samples and post-diagnosis urine samples from cancer patients. These strengths allow us to evaluate study hypotheses more rigorously than many existing cohort studies are able to do. The SMHS has made important contributions to NCD research and its value will continue to grow with extended follow-up.

Can I get access to the data? Where can I find out more?

Permission is required to access SMHS data and resources. The SMHS survey questionnaires, publications, policy for data sharing and request procedures can be found at: http://www.mc.vanderbilt.edu/swhs-smhs/. Questions for potential collaboration can also been sent to Dr Xiao-Ou Shu at: [xiao-ou.shu@vanderbilt.edu].

Acknowledgements

The authors thank participants and research staff members of the Shanghai Men's Health Study for their dedication to the study, and Nan Kennedy for editing and manuscript preparation.

Funding

This work was supported by the US National Cancer Institute at the National Institutes of Health [grant numbers R01 CA082729, UM1 CA173640], and partially by the State Key Project Specialized for Infectious Diseases, China [grant numbers 2008ZX10002-015, 2012ZX10002008-002].

Conflict of interest: None declared.

References

1. World Health Organization. *Noncommunicable Diseases, Country Profiles 2014*. Geneva: World Health Organization, 2014.
2. World Health Organization. *Projections of Mortality and Causes of Death, 2015–2030*. Geneva: World Health Organization, 2014.
3. Bell AC, Ge K, Popkin BM. Weight gain and its predictors in Chinese adults. *Int J Obes Relat Metab Disord* 2001;**25**:1079–86.
4. Jia WP, Pang C, Chen L *et al*. Epidemiological characteristics of diabetes mellitus and impaired glucose regulation in a Chinese adult population: the Shanghai Diabetes Studies, a cross-sectional 3-year follow-up study in Shanghai urban communities. *Diabetologia* 2007;**50**:286–92.
5. Gu D, Gupta A, Muntner P *et al*. Prevalence of cardiovascular disease risk factor clustering among the adult population of China: results from the International Collaborative Study of Cardiovascular Disease in Asia (InterAsia). *Circulation* 2005;**112**:658–65.
6. Gao YT, Lu W (eds). *Cancer Incidence, Mortality and Survival Rates in Urban Shanghai (1973-2000)*. Shanghai, China: Second Military Medical University Press, 2007.
7. Yang G, Kong L, Zhao W *et al*. Emergence of chronic non-communicable diseases in China. *Lancet* 2008;**372**:1697–705.
8. Villegas R, Yang G, Liu D *et al*. Validity and reproducibility of the food-frequency questionnaire used in the Shanghai Men's Health Study. *Br J Nutr* 2007;**97**:993–1000.
9. Jurj AL, Wen W, Xiang Y-B *et al*. Reproducibility and validity of the Shanghai Men's Health Study physical activity questionnaire. *Am J Epidemiol* 2007;**165**:1124–33.
10. Lee S-A, Wen W, Xiang Y-B *et al*. Assessment of dietary isoflavone intake among middle-aged Chinese men. *J Nutr* 2007;**137**:1011–16.
11. Zhang W, Shu X-O, Li H *et al*. Vitamin intake and liver cancer risk: a report from two cohort studies in China. *J Natl Cancer Inst* 2012;**104**:1173–81.
12. Yang G, Zheng W, Xiang Y-B *et al*. Green tea consumption and colorectal cancer risk: a report from the Shanghai Men's Health Study. *Carcinogenesis* 2011;**32**:1684–88.
13. Epplein M, Shu XO, Xiang YB *et al*. Fruit and vegetable consumption and risk of distal gastric cancer in the Shanghai Women's and Men's Health studies. *Am J Epidemiol* 2010;**172**:397–406.
14. Villegas R, Xiang Y-B, Elasy T *et al*. Fish, shellfish, and long-chain n-3 fatty acid consumption and risk of incident type 2 diabetes in middle-aged Chinese men and women. *Am J Clin Nutr* 2011;**94**:543–51.
15. Zhang X, Shu X-O, Xiang Y-B *et al*. Cruciferous vegetable consumption is associated with a reduced risk of total and cardiovascular disease mortality. *Am J Clin Nutr* 2011;**94**:240–46.
16. Takata Y, Zhang X, Li H *et al*. Fish intake and risks of total and cause-specific mortality in 2 population-based cohort studies of 134,296 men and women. *Am J Epidemiol* 2013;**178**:46–57.
17. Wang N, Zhang X, Xiang Y-B *et al*. Associations of tai chi, walking, and jogging with mortality in Chinese men. *Am J Epidemiol* 2013;**178**:791–96.
18. Epplein M, Zheng W, Xiang Y-B *et al*. Prospective study of helicobacter pylori biomarkers for gastric cancer risk among Chinese men. *Cancer Epidemiol Biomarkers Prev* 2012;**21**:2185–92.

19. Wu J, Cai Q, Li H *et al*. Circulating C-reactive protein and colorectal cancer risk: a report from the Shanghai Men's Health Study. *Carcinogenesis* 2013;**34**(12):2799–803.
20. Takata Y, Shrubsole MJ, Li H *et al*. Plasma folate concentrations and colorectal cancer risk: A case-control study nested within the Shanghai Men's Health Study. *Int J Cancer* 2014;**135**(9): 2191–98.
21. Jia W-H, Zhang B, Matsuo K *et al*. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet* 2013;**45**(2):191–96.
22. Zhang B, Jia W-H, Matsuda K *et al*. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* 2014;**46**(6):533–42.
23. Abnet CC, Freedman ND, Hu N *et al*. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* 2010;**42**(9): 764–67.
24. Wen W, Cho Y-S, Zheng W *et al*. Meta-analysis identifies common variants associated with body mass index in East Asians. *Nat Genet* 2012;**44**(3):307–11.
25. Wen W, Zheng W, Okada Y *et al*. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet* 2014; **23**(20):5492–504.
26. Zhang B, Jia W-H, Matsuo K *et al*. Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. *Int J Cancer* 2014;**135**:948–55.
27. Zheng W, McLerran DF, Rolland B *et al*. Association between body-mass index and risk of death in more than 1 million Asians. *N Engl J Med* 2011;**364**:719–29.
28. Zheng W, McLerran DF, Rolland BA *et al*. Burden of total and cause-specific mortality related to tobacco smoking among adults aged $\geq$ 45 years in Asia: a pooled analysis of 21 cohorts. *PLoS Med* 2014;**11**:e1001631.
29. Chen Y, Copeland WK, Vedanthan R *et al*. Association between body mass index and cardiovascular disease mortality in east Asians and south Asians: pooled analysis of prospective data from the Asia Cohort Consortium. *BMJ* 2013;**347**:f5446.
30. King H, Li JY, Locke FB, Pollack ES, Tu JT. Patterns of site-specific displacement in cancer mortality among migrants: the Chinese in the United States. *Am J Public Health* 1985;**75**: 237–42.
31. Chan JCN, Malik V, Jia W *et al*. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *JAMA* 2009;**301**:2129–40.
32. Adlercreutz H, Mazur W. Phyto-oestrogens and Western diseases. *Ann Med* 1997;**29**:95–120.
33. D'Adamo CR, Sahin A. Soy foods and supplementation: a review of commonly perceived health benefits and risks. *Altern Ther Health Med* 2014;**20**(**Suppl 1**):39–51.
34. Hecht SS. Chemoprevention of cancer by isothiocyanates, modifiers of carcinogen metabolism. *J Nutr* 1999;**129**:768S–74S.
35. Eilat-Adar S, Sinai T, Yosefy C, Henkin Y. Nutritional recommendations for cardiovascular disease prevention. *Nutrients* 2013;**5**:3646–83.
36. Curado MP, Edwards B, Shin HR *et al*. (eds). *Cancer Incidence in Five Continents*. Vol IX. Lyon, France: IARC Scientific Publications, 2009.
37. Jemal A, Tiwari RC, Murray T *et al*. Cancer statistics, 2004. *CA Cancer J Clin* 2004;**54**:8–29.
38. *Shanghai Cancer Report 2010*. Shanghai, China: Shanghai Municipal Center for Disease Control & Prevention, 2010.
39. Ziegler RG, Hoover RN, Pike MC *et al*. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst* 1993;**85**:1819–27.
40. Dai Q, Franke AA, Yu H *et al*. Urinary phytoestrogen excretion and breast cancer risk: evaluating potential effect modifiers, endogenous estrogens and anthropometrics. *Cancer Epidemiol Biomarkers Prev* 2003;**12**:497–502.
41. Dai Q, Gao Y-T, Shu X-O *et al*. Oxidative stress, obesity, and breast cancer risk: results from the Shanghai Women's Health Study. *J Clin Oncol* 2009;**27**:2482–88.
42. Cai Q, Gao Y-T, Chow W-H *et al*. Prospective study of urinary prostaglandin E2 metabolite and colorectal cancer risk. *J Clin Oncol* 2006;**24**:5010–16.
43. Fowke JH, Chung F-L, Jin F *et al*. Urinary isothiocyanate levels, brassica, and human breast cancer. *Cancer Res* 2003;**63**: 3980–86.
44. Zheng W, Jin F, Dunning LA *et al*. Epidemiological study of urinary 6-beta-hydroxycortisol to cortisol ratios and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2001;**10**: 237–42.
45. Wu Q-J, Wang J, Gao J *et al*. Urinary isothiocyanates level and liver cancer risk: a nested case-control study in Shanghai, China. *Nutr Cancer* 2014;**66**:1023–29.
46. Zheng W, Dai Q, Custer LJ *et al*. Urinary excretion of isoflavonoids and the risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 1999;**8**:35–40.
47. Takata Y, Xiang Y-B, Yang G *et al*. Intakes of fruits, vegetables, and related vitamins and lung cancer risk: results from the Shanghai Men's Health Study (2002-2009). *Nutr Cancer* 2013; **65**:51–61.
48. Luu HN, Blot WJ, Xiang Y-B *et al*. Prospective evaluation of the association of nuts/peanut consumption with total and cause-specific mortality. *JAMA Intern Med* 2015;**175**:755–66.
49. Takata Y, Shu X-O, Gao Y-T *et al*. Red meat and poultry intakes and risk of total and cause-specific mortality: results from cohort studies of Chinese adults in Shanghai. *PloS One* 2013;**8**: e56963.
50. Yu D, Zhang X, Xiang Y-B *et al*. Association of soy food intake with risk and biomarkers of coronary heart disease in Chinese men. *Int J Cardiol* 2014;**172**:e285–87.
51. Zhang X, Shu X-O, Li H *et al*. Visceral adiposity and risk of coronary heart disease in relatively lean Chinese adults. *Int J Cardiol* 2013;**168**:2141–45.
52. Wang N, Zhang X, Xiang Y-B *et al*. Associations of adult height and its components with mortality: a report from cohort studies of 135,000 Chinese women and men. *Int J Epidemiol* 2011;**40**: 1715–26.
53. Shu XO, Long J, Cai Q *et al*. Identification of new genetics risk variants for Type 2 diabetes. *PloS Genet* 2010;**6**:pii:e1001127.
54. Cho YS, Chen C-H, Hu C *et al*. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 2012;**44**:67–72.