Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2013;42:731–740
doi:10.1093/ije/dys053

COHORT PROFILE

Cohort Profile: The Dongfeng–Tongji cohort study of retired workers

Feng Wang,[1] Jiang Zhu,[2,3] Ping Yao,[1] Xiulou Li,[2,3] Meian He,[1] Yuewei Liu,[1] Jing Yuan,[1] Weihong Chen,[1] Li Zhou,[1] Xinwen Min,[2,3] Weimin Fang,[1] Yuan Liang,[1] Youjie Wang,[1] Sheng Wei,[1] Jichun Liu,[4] Xiaoping Miao,[1] Mingjian Lang,[2,3] Xixiang Jiang,[4] Peng Zhang,[2,3] Dongfeng Li,[2,3] Chuanwen Lu,[4] Xiaozheng Wang,[2,3] Wenhu Shi,[4] Jianru Zheng,[2,3] Huan Guo,[1] Xiaomin Zhang,[1] Handong Yang,[2,3] Frank B Hu[5,6] and Tangchun Wu[1]*

[1]MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, Hubei, China, [2]Dongfeng Central Hospital, Dongfeng Motor Corporation, Shiyan 442008, Hubei, China, [3]Hubei University of Medicine, Shiyan 442008, Hubei, China, [4]Social Insurance Center, Dongfeng Motor Corporation, Shiyan 442008, Hubei, China, [5]Department of Nutrition, Harvard School of Public Health, Boston, 02115 MA, USA and [6]Department of Epidemiology, Harvard School of Public Health, Boston, 02115 MA, USA

*Corresponding author. MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, Hubei, China. E-mail: wut@mails.tjmu.edu.cn

Accepted 13 March 2012

How did the study come about?

China has seen rapid socio-economic and epidemiological changes over the past several decades. Economic growth plus shifts in environment, lifestyles and diet have increased life expectancy, but they have also led to a higher burden of chronic, non-communicable diseases. Stroke, coronary heart disease (CHD), cancer and diabetes account for 80% of the deaths and 70% of the disability-adjusted life-years lost in China.[1,2] In recent decades, epidemics of obesity and metabolic syndrome have grown rapidly in China, although the prevalence is still relatively low compared with Western countries.[3–7]

Multiple factors have contributed to China's current epidemic of chronic diseases, including reduced infectious illnesses, a rapidly ageing population, adoption of a Western diet and lifestyle and other risk factors, such as smoking, excessive alcohol use, pollution and psychosocial stress.[8] These trends, which are accelerating during a period of unprecedented urbanization, are having an increasingly negative effect on public health in China.

The growing burden of chronic disease may be due in part to unique environmental and occupational exposures as well as genetic factors in the Chinese population. However, their roles have yet to be clearly determined. Evidence suggests that Asians, including the Chinese, are more susceptible to diabetes, hypertension and other chronic diseases than Caucasians for a given age and body mass index (BMI).[9,10] Recent advances in genome-wide association studies (GWAS) have led to successful identification of a large number of common variants for chronic diseases, including obesity, diabetes, cardiovascular diseases and cancer.[11] However, few studies have examined potential interactions between genetic and environmental risk factors.[12] Rapid epidemiological transitions in China provide a unique opportunity to examine a wide range of factors behind the causes and progression of chronic diseases and the potential role of gene–environment interactions.

To investigate these factors, we launched the Dongfeng–Tongji cohort (DFTJ cohort) study in Shiyan City, Hubei, China. The Dongfeng Motor Corporation (DMC), founded in 1969, is one of the three largest auto manufacturers in China. DMC is a state-owned enterprise, which is located in Shiyan City, Hubei province (Figure 1). Most first-generation employees, almost all retired now, emigrated to Hubei province from many parts of China, such as Liaoning, Jilin, Shandong, Jiangsu and Shanghai. All retired employees are covered by DMC's health-care service system, which consists of five company-owned hospitals, one Center of Disease Control and Prevention (CDC) and one Social Insurance Center. Dongfeng Central Hospital is the largest medical centre in the system and provides comprehensive care for all

Figure 1 Locations of DMC in China

employees including retired employees. The CDC monitors occupational hazards and diagnoses of work-related diseases. The medical insurance system plays a role in disease registry and management. Retired employees have ready access to the system for general health check-ups and medical care.

The DFTJ cohort study was initiated by Tangchun Wu (School of Public Health, Tongji Medical College, HUST), Handong Yang (Dongfeng Central Hospital, DMC), Jichun Liu (Social Insurance Center, DMC) and Frank Hu (Harvard School of Public Health, Boston, MA, USA). It is a long-term prospective cohort study of DMC employees. Funding for the recruitment, examination and follow-up of participants was obtained from the Huazhong University of Science and Technology, the Medical Insurance Center and the Dongfeng Central Hospital, DMC.

The main goal of the study is to examine the determinants of obesity, diabetes, metabolic syndrome, cardiovascular disease and other chronic diseases in a population that is in the midst of rapid epidemiological transition. There is growing evidence that the gene–environment interactions play an important role in the development of chronic diseases. For example, the HFE H63D polymorphism has been shown to modify the effects of the cumulative lead exposure on pulse pressure, which is a predictor of cardiovascular disease.[13] Also, occupational exposure to polycyclic aromatic hydrocarbons was found to interact with the GSTP1 IIe105Val polymorphism on risk of developing prostate cancer.[14] In addition, specific genetic polymorphisms were found to modify the effects of noise on hearing loss,[15] the effects of benzene on aplastic anaemia and leukaemia[16] and the effects of phthalate on leiomyoma.[17] In the present study, we have collected occupational exposure factors including noise, vibration, dust, high temperature and organic toxins each year by experienced occupational health staff since 1993. The exposure data are available among 79.3% of the participants and the mean exposure time was 6.7 years. Therefore, it is feasible for us to identify the environmental and genetic risk factors for a wide range of chronic diseases and to investigate gene–environment interactions and novel biomarkers in the prediction of chronic disease incidence and mortality. Also, because participants included in this study are middle-aged and older adults, higher prevalence and incidence rates of the chronic diseases enable us to have enough cases to investigate the risk factors including genetic and environmental factors and their interaction on these diseases. Ultimately, this project will help advance our understanding of the causes of chronic diseases and will aid in the development of strategies to prevent and control them in China.

What does the study cover?

The study investigates a wide range of lifestyle, dietary, psychosocial, occupational and environmental factors and biochemical and genetic factors in relation to the development of chronic diseases. Towards this end, we have designed a standard questionnaire that covers diet, lifestyle, occupational history, environmental exposures and medical history. We will track occupational exposure through workplace monitoring at DMC for the retired employees enrolled in the present cohort study and obtain environmental exposure information from the Environmental Protection Agency of Shiyan City, Hubei.

In the short-term follow-up, we focus on the role of genetic and environmental factors in the development of obesity, metabolic syndrome and type 2 diabetes. With long-term follow-up, we will be able to study cardiovascular and cerebrovascular outcomes and cancer. Classical cohort studies such as the Framingham Heart Study,[18] the Nurses' Health Study[19] and the British Doctors' Study[20] have investigated a wide range of chronic disease outcomes such as cancer and cardiovascular diseases as well as total and cause-specific mortality. This cohort will give us the opportunity to do likewise in a Chinese population. Archived blood and DNA samples will be used for nested case–control analyses of biochemical and genetic markers and future risk of chronic diseases. We will also analyse biospecimens using emerging technologies, such as metabolomics, transcriptomics and epigenetics.

Who is in the sample?

The Retirement Office and the Social Insurance Center at DMC provided a list of the retired employees of the company and invited them to participate in this study. A total of 31 000 retired employees were invited at DMC. For those who did not respond to the invitation, the staff at the Retirement Office followed them up with a telephone call. Approximately 87% ($n = 27\,009$) of the invited participants agreed and provided baseline blood samples and questionnaire information between September 2008 and June 2010. Among those who did not participate in this study, most ($n = 3000$) moved to other cities. The socio-demographic characteristics were similar between the responders and non-responders (Table 1).

After obtaining written informed consent, trained interviewers use a semi-structured questionnaire to collect baseline data during face-to-face interviews. The general health examination was performed at the same time. Figure 2 shows the overall study plan. Exclusion criteria included chemotherapy or radiotherapy for cancer and other severe diseases. The Medical Ethics Committee of the School of Public Health, Tongji Medical College, HUST and Dongfeng General Hospital, DMC, approved this study.

How often will participants be followed up and what is the likely rate of loss?

DMC medical insurance covers the cost of primary care received within the system. This allows us to track health-service use, disease incidence and mortality through this medical insurance system and Dongfeng Central hospital. We will contact participants every 5 years to repeat the questionnaire interview and physical examination (Figure 2). Self-reported chronic diseases, such as CHD, diabetes, stroke and major cancers, are verified through medical record reviews. The diagnosis of these conditions is based on well-accepted international standards.[21–23] Electronic medical records in the DMC hospitals allow us to link inpatient and outpatient records to our database on an ongoing basis. These data enable disease follow-up and documentation of deaths.

Each participant has a unique medical insurance card number and ID, making it easy to track disease incidence and collect biosamples of patients with existing or newly occurring diseases. Causes of death are classified according to the 10th version of the International Statistical Classification of Diseases

Table 1 A comparison of the socio-demographic characteristics between the responders and non-responders

Variables	Responders (27 009)	Non-responders (3991)	*P*-value
Male, *n* (%)	12 052 (44.6)	2175 (54.5)	<0.001
Age (years), *n* (%)	63.59 (7.83)	63.39 (8.98)	0.20
<50	722 (2.7)	212 (5.5)	
−50	2669 (9.9)	392 (9.9)	
−55	4477 (16.7)	622 (15.7)	
−60	7754 (28.9)	1097 (27.7)	
−65	5204 (19.4)	614 (15.5)	
−70	3542 (13.2)	548 (13.8)	
⩾75	2465 (9.2)	480 (12.1)	
Race, *n*(%)			0.22
Han Chinese	26 530 (98.2)	3900 (97.7)	
Other ethnic groups	395 (1.5)	48 (1.2)	
Missing	84 (0.3)	43 (1.1)	
Education, *n*(%)			0.10
Primary school or illiteracy	7847 (29.3)	1086 (27.4)	
Middle school	9618 (35.9)	1480 (37.4)	
High school	6478 (24.2)	965 (24.4)	
University or college or higher	2844 (10.6)	426 (10.7)	
Marital status, *n* (%)			<0.001
Single	79 (0.3)	23 (0.6)	
Married	24 150 (89.7)	3506 (88.2)	
Widowed	2183 (8.1)	339 (8.5)	
Divorced	525 (1.9)	108 (2.7)	

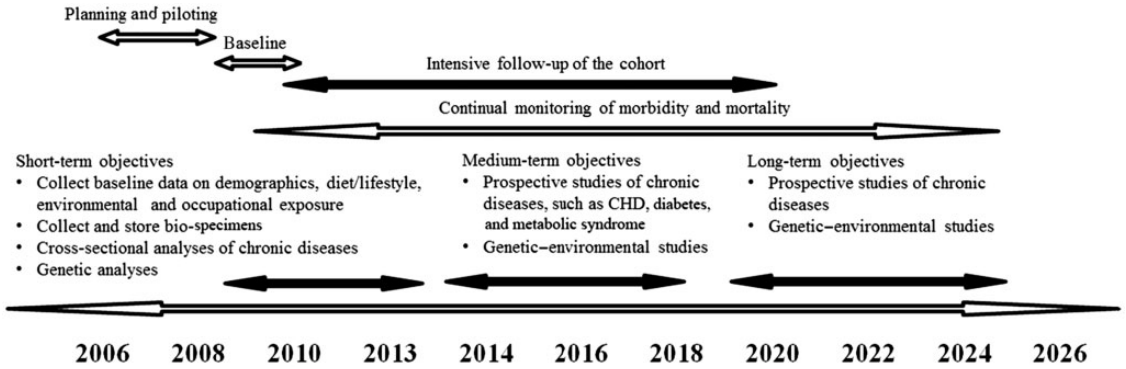

Figure 2 Study plan of the DFTJ cohort

Table 2 Summary of clinical measures collected at baseline in the DFTJ cohort

Variables	Number of measurements	Equipment used
Standing height	Once	Dekon DK-08-E, Rayweigh, Beijing, China
Body weight	Once	Dekon DK-08-E, Rayweigh Beijing, China
Waist circumference	Twice	Hoechstmass, Germany
Resting blood pressure	Thrice	Mercury sphygmomanometer, Shanghai Zhangdong Med-Tech Ltd Company, China
12-lead electrocardiography	Once	Cardiofax ECG-9020P, NIHO KHDEN, China
Chest X-ray	Once	Shimadzu UD150L, Shimadzu, Japan
Abdominal B-type ultrasound inspection	Once	Aplio XG, TOSHIBA, Japan
Fasting plasma glucose	Once	ARCHITECT ci8200, Abbott, USA
Blood lipids	Once	ARCHITECT ci8200, Abbott, USA
Hepatic function	Once	ARCHITECT ci8200, Abbott, USA
Renal function	Once	ARCHITECT ci8200, Abbott, USA
Complete blood count	Once	CELL-DYN 3700, Abbott, USA
Tumour biomarkers	Once	ARCHITECT ci8200, Abbott, USA
Urine routine	Once	Mejer-600, Mejer, China

and Related Health Problems. Clinical data are derived from the Medical Insurance Center records and the Dongfeng Central Hospital, DMC. The system enables us to locate participants who move to other cities and will minimize the expected rate of loss to follow-up to less than an estimated 10%.

What is being measured?

After an overnight fast, all participants came to the health examination centre at Dongfeng Central Hospital, where trained physicians, nurses and technicians performed physical examinations. The physical component measures included standing height, body weight, waist circumference and blood pressure. Clinical examinations were conducted for potential conditions in the liver, gall bladder, spleen, kidney, prostate (for males) and the uterus, ovaries and fallopian tubes (for females) (Table 2). Standing height, body weight and waist circumference were measured in participants with light indoor clothing and without shoes.

Trained interviewers administered questionnaires during face-to-face interviews. The questionnaires included demographic information, occupational history, socio-economic status, family and personal disease histories, smoking history, alcohol use, diet (determined via a simplified semi-quantitative food frequency questionnaire), physical activity, stress and psychological status (Table 3). In addition, environmental exposure information was obtained from the Environmental Protection Agency of Shiyan City, Hubei, and the occupational exposure (such as noise, vibration, dust, high temperature, organic toxins and so on) of the retired employees was tracked through the workplace monitoring at DMC. The assessment of occupational exposures was carried out by the experienced occupational health staff once a year according to the national occupational health criteria of China

Table 3 Summary of questionnaire data collected at baseline from the DFTJ cohort

Exposure category	Variable/exposure
Demographics and socio-economics	Birthday, race, religion, housing type, household size, income and education
Personal health behaviour	Smoking, alcohol, tea, afternoon nap and physical activity (at work and leisure)
Diet	Major food groups, including meats, vegetables, fruits, beans, eggs and dairy
Environmental exposure	Occupational history, living environment and passive smoking exposure
Family history	Family history of hypertension, hyperlipidemia, CHD, diabetes, stroke, cancers, emphysema, chronic bronchitis, asthma, pulmonary tuberculosis, cholelithiasis, chronic hepatitis, nephritis and arthritis
Past medical history	Diagnosed medical conditions, use of health services and use of medicines for the most recent 2 weeks
Reproductive history	Parity and breastfeeding history, menopause status and associated timing, contraceptive history, history of hormone replacement therapy use and gynaecological diseases
Stress and psychological status	Psychological morbidity scales/stress, anger/hostility, optimism and social isolation

GBZ/T 189.8-2007 for the measurement of noise, GB 10434-1989 for vibration, GBZ/T192-2007 for the dust and GB/T934-2008 for the high temperature in workplace.

Fifteen millilitres of fasting blood was drawn with three vacuum tubes [two ethylenediamine tetraacetic acid (EDTA) anticoagulation tubes for plasma and DNA and one coagulation tube for serum]. The hospital's laboratory measured blood lipids [total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C)], fasting glucose, hepatic function [bilirubin, alanine aminotransferase (ALT), aspartate aminotransferase (AST) and alkaline phosphatase (ALP)] and renal function (blood urea nitrogen (BUN), creatinine and uric acid].

The laboratory also provided a complete blood count [including red blood cell count (RBC); mean corpuscular volume (MCV); haemoglobin (Hb); haematocrit (Hct); mean corpuscular haemoglobin (MCH); mean corpuscular haemoglobin concentration (MCHC); erythrocyte haemoglobin distribution width (RDW); white blood cell count (WBC); count and percentage of eosinophil granulocytes, basophil granulocytes, neutrophil granulocytes, monocytes and lymphocytes; platelet count (PLT) and mean platelet volume (MPV); thrombocytocrit (THR); platelet volume distribution width (PDW)] and urine routine test (urine pH, nitrite, glucose, vitamin C, white blood cell, urine protein, bilirubin, urobilinogen, urine ketobody, urine-specific gravity and occult blood).

Tumour-associated antigens [cancer antigen 125 (CA125), squamous cell carcinoma-associated antigen (SCC), cancer antigen 19-9 (CA19-9), carcinoembryonic antigen (CEA) and alpha fetoprotein (AFP)] were determined by immunoassay at the same laboratory. Aliquots of plasma, serum, whole blood cells and DNA were stored at −80°C. All biospecimens will be tracked through an electronic database. In particular, a GWAS was conducted to investigate genetic determinants of the biochemical traits mentioned above in a subsample of the participants ($n = 1461$). Validation of the genome-wide significant variants was completed in more than 8000 participants. In addition, four variants associated with the uric acid have been genotyped in more than 10 000 participants to investigate the association of these variants with metabolic syndrome and explore potential gene–environment interactions.

Trained investigators entered questionnaire data into the computer twice using EpiData software. The SAS software was used to establish the DFTJ cohort study database, which links data from multiple sources (including questionnaire data, laboratory data and clinical examination data).

What has been found?

Recruitment of participants took place during September 2008 to June 2010; 27 009 retired individuals were examined and the response rate was ∼87%. Males and those who were divorced tended to have a higher non-response rate. The distributions of other variables were similar between the responders and the non-responders (Table 1). Among the responders, 98.2% were Han Chinese. The mean age was 63.6 years (66.3 years for males, 61.4 years for females). Females made up 55.4% of the sample. Among males, 37.0% were current smokers and 25.2% former smokers. Among females, 2.4% were current smokers and 0.9% former smokers. Current alcohol use rates were 39.0% and 6.2% for males and females, respectively. The mean weight was 63.5 kg (68.0 kg for males, 59.9 kg for females) and

Table 4 The baseline characteristics of the participants in the DFTJ cohort

Variables	Male, *n* (%)	Female, *n* (%)	Total, *n* (%)
Gender	12 052 (44.6)	14 957 (55.4)	27 009 (100.0)
Age (years)	66.33 (6.70)	61.40 (7.98)	63.59 (7.83)
<50	220 (1.8)	502 (3.4)	722 (2.7)
−50	142 (1.2)	2527 (17.0)	2669 (9.9)
−55	708 (5.9)	3769 (25.4)	4477 (16.7)
−60	4360 (36.4)	3394 (22.8)	7754 (28.9)
−65	3163 (26.4)	2041 (13.7)	5204 (19.4)
−70	1933 (16.2)	1609 (10.8)	3542 (13.2)
⩾75	1442 (12.0)	1023 (6.9)	2465 (9.2)
Race			
Han Chinese	11 817 (98.1)	14 713 (98.4)	26 530 (98.2)
Other ethnic groups	180 (1.5)	215 (1.4)	395 (1.5)
Missing	55 (0.5)	29 (0.2)	84 (0.3)
Weight (kg)	68.0 (10.2)	59.9 (9.2)	63.5 (10.4)
Height (cm)	166.3 (6.1)	156.3 (5.8)	160.8 (7.7)
BMI	24.6 (3.2)	24.5 (3.5)	24.5 (3.4)
Waist circumference (cm)	85.3 (9.3)	81.6 (9.4)	83.2 (9.5)
Systolic blood pressure (mmHg)	131.4 (18.4)	128.7 (19.0)	130.0 (18.8)
Diastolic blood pressure (mmHg)	78.5 (11.1)	77.5 (10.7)	77.9 (10.9)
Physical activity			
Yes	10 778 (89.6)	13 132 (88.0)	23 910 (88.7)
No	1247 (10.4)	1794 (12.0)	3041 (11.3)
Smoking			
Current smoker	4442 (37.0)	355 (2.4)	4797 (17.9)
Former smoker	3024 (25.2)	140 (0.9)	3164 (11.8)
Non-smoker	4532 (37.8)	14 816 (96.7)	18 853 (70.3)
Drinking			
Current drinker	4702 (39.0)	922 (6.2)	5624 (20.8)
Former drinker	1389 (11.5)	164 (1.1)	1553 (5.8)
Non-drinker	5950 (49.4)	13 853 (92.7)	19 803 (73.4)
Education			
Primary school or illiteracy	3400 (28.4)	4447 (30.0)	7847 (29.3)
Middle school	4151 (34.7)	5467 (36.9)	9618 (35.9)
High school	2608 (21.8)	3870 (26.1)	6478 (24.2)
University or college or higher	1811 (15.2)	1033 (7.0)	2844 (10.6)
Marital status			
Single	40 (0.3)	39 (0.3)	79 (0.3)
Married	11 325 (94.2)	12 825 (86.0)	24 150 (89.7)
Widowed	509 (4.2)	1674 (11.2)	2183 (8.1)
Divorced	147 (1.2)	378 (2.5)	525 (1.9)
Shift work	4300 (35.7)	5109 (34.2)	9409 (34.8)
Branch factory			
Auto raw material manufacture	781 (6.5)	1230 (8.2)	2011 (7.4)
Auto parts manufacture	2328 (19.3)	5312 (35.5)	7640 (28.3)

(continued)

Table 4 Continued

Variables	Male, *n* (%)	Female, *n* (%)	Total, *n* (%)
Auto assembly manufacture	202 (1.7)	360 (2.4)	562 (2.1)
Auto affiliated manufacture	847 (7.0)	1991 (13.3)	2838 (10.6)
Management and service organization	4284 (35.6)	2381 (16.0)	6665 (24.7)
Mechanical manufacture	2300 (19.1)	1499 (10.0)	3799 (14.1)
Others	1310 (10.9)	2184 (14.6)	3494 (12.9)
Exposed to occupational factors			
Noise	1472 (11.8)	1897 (12.7)	3369 (12.5)
Vibration	416 (3.3)	730 (4.9)	1146 (4.2)
Dust	1288 (10.3)	1751 (11.7)	3039 (11.3)
High temperature	506 (4.2)	424 (2.8)	930 (3.4)
Organic toxins	1727 (13.8)	2406 (16.1)	4133 (15.3)
Medication use in recent 2 weeks			
Yes	6673 (55.5)	7899 (52.9)	14 572 (54.0)
No	5357 (44.5)	7040 (47.4)	12 397 (46.0)
Hormone replacement therapy			
Yes	–	456 (3.2)	–
No	–	14 230 (96.8)	–
Hypertension	4880 (40.5)	5398 (36.1)	10 278 (38.1)
Hyperlipidemia	2869 (23.9)	3201 (21.4)	6070 (22.5)
Diabetes mellitus	1396 (11.6)	1801 (12.0)	3197 (11.8)
CHD	2008 (16.7)	2110 (14.1)	4118 (15.2)
Stroke	735 (6.1)	393 (2.6)	1128 (4.2)
Myocardial infraction	474 (3.9)	257 (1.7)	731 (2.7)
Cancer	455 (3.8)	1008 (6.7)	1463 (5.4)
Emphysema	392 (3.3)	147 (1.0)	539 (2.0)
Chronic bronchitis	1761 (14.6)	1294 (8.7)	3065 (11.3)
Asthma	522 (4.3)	372 (2.5)	894 (3.3)
Pulmonary tuberculosis	582 (4.8)	358 (2.4)	940 (3.5)
Cholelithiasis	1193 (9.9)	2054 (13,7)	3247 (12.0)
Chronic hepatitis	681 (5.7)	313 (2.1)	994 (3.7)
Nephritis	386 (3.2)	633 (4.2)	1019 (3.8)

mean height was 160.8 cm (166.3 cm for males, 156.3 cm for females). The mean waist circumference was 83.2 cm (85.3 cm for males and 81.6 cm for females). The mean systolic blood pressure and diastolic blood pressure was 130.0 and 77.9 mmHg, respectively. Among the participants, 35.9% graduated from middle school and 10.6% obtained college or higher education (15.2% for male and 7.0% for female) and 34.8% participants had shift work experience. The proportion of workers exposed to noise was 11.8% for males and 12.7% for females, the vibration exposure was 3.3 and 4.9%, the dust exposure was 10.3 and 11.7%, the high temperature exposure was 4.2 and 2.8% and the organic toxin exposure was 13.8 and 16.1% for males and females, respectively. Most of the participants (89.7%) were married. In females, only 3.2% took hormone replacement therapy (Table 4).

The mean TC, TG, HDL-C and LDL-C was 5.2, 1.5, 1.4 and 3.0 mmol/l, respectively. The mean values and distributions of hepatic function parameters, lipids, fasting glucose, renal function parameters and haematological trait are shown in Table 5.

The prevalence of hypertension was 40.5% for males and 36.1% for females. The prevalence of self-reported stroke was 2-fold higher in males than in females, whereas the prevalence of self-reported CHD and diabetes was similar between males and females (Table 4). We are in the process of confirming these self-reported diseases through a review of medical

Table 5 The baseline levels of biochemical traits of the participants in the DFTJ cohort (mean ± SD)

Variables	Male	Female	Total
Hepatic function			
ALP (U/l)	88.0 (29.1)	93.9 (38.7)	91.3 (34.9)
AST (U/l)	26.4 (17.4)	24.7 (13.0)	25.4 (15.1)
ALT (U/l)	25.5 (22.5)	23.3 (17.6)	24.3 (20.0)
Total bilirubin (μmol/l)	15.5 (6.2)	13.0 (5.1)	14.1 (5.8)
Direct bilirubin (μmol/l)	4.5 (2.1)	3.6 (2.0)	4.0 (2.1)
Indirect bilirubin (μmol/l)	11.0 (5.0)	9.4 (4.3)	10.1 (4.7)
Lipids			
TC (mmol/l)	5.0 (0.9)	5.3 (1.0)	5.2 (1.0)
TG (mmol/l)	1.4 (1.2)	1.5 (1.1)	1.5 (1.1)
HDL-C (mmol/l)	1.4 (0.4)	1.5 (0.4)	1.4 (0.4)
LDL-C (mmol/l)	2.9 (0.8)	3.1 (0.8)	3.0 (0.8)
Fasting glucose (mmol/l)	6.2 (1.8)	6.0 (1.7)	6.1 (1.7)
Renal function			
BUN (mmol/l)	5.6 (2.4)	5.1 (1.8)	5.3 (2.1)
Creatinine (μmol/l)	90.5 (33.6)	72.6 (20.2)	80.7 (28.5)
Uric acid (μmol/l)	328.6 (81.8)	265.1 (73.0)	293.8 (83.3)
Haematological traits			
RBC (t/l)	4.8 (0.5)	4.4 (0.4)	4.6 (0.5)
WBC (g/l)	6.3 (1.7)	5.8 (1.6)	6.1 (1.7)
Platelet count (g/l)	178.1 (54.8)	194.3 (57.6)	186.9 (56.9)
Haemoglobin (g/l)	144.8 (13.5)	129.7 (11.3)	136.6 (14.4)

records. Also, a GWAS (using Affymetrix Array 6.0 chips) was conducted to investigate genetic determinants of biochemical traits including lipids (TC, TG, HDL-C and LDL-C), liver function traits (bilirubin, AST, ALT and ALP), kidney function traits (BUN, creatinine and uric acid) and tumour biomarkers (CA125, SCC, CA19-9, CEA and AFP) in a subsample of the participants ($n = 1461$). For those SNPs reaching the genome-wide significant threshold, a validation study was conducted in more than 8000 participants. The results of the GWAS will be reported in separate articles. We plan to conduct additional genotyping of common and rare variants in a larger number of participants aiming to identify genetic risk factors for metabolic diseases and to explore gene–environment interactions.

What are the main strengths and weaknesses?

This study has several clear strengths. First, the large sample size will provide ample power for the analyses of genetic and environmental determinants of chronic diseases, such as metabolic syndrome, diabetes, CHD, stroke and pulmonary diseases. Secondly, the study population is unique in its occupational and environmental exposures, and these data have been collected from regular monitoring of occupational exposure from DMC since 1993. We have also collected detailed information on demographics, diet and lifestyle and psychosocial variables at baseline. Thirdly, disease diagnoses and follow-up are linked to our cohort database through the health service system's medical records, enabling real-time confirmation of patient health status and tracking of medical histories, prescriptions and clinical parameters. Fourthly, DMC's full coverage of medical care and insurance will minimize loss to follow-up. Finally, the large number of archived biospecimens will allow us to conduct nested case–control studies of biochemical and genetic markers of chronic diseases. The power for analysing cardiovascular outcomes and individual cancers will be enhanced over time.

A main weakness of this cohort study is that the participants are all middle-aged and older adults and we do not have information on early-life exposures. There is increasing evidence that early-life exposures may have long-term influences on chronic disease risk in adult life,[24] and a life course approach would be more powerful conceptually. However, the high prevalence and incidence of chronic diseases in

middle-aged and older adults in the context of rapid epidemiological transitions will provide us with sufficient power for the proposed analyses. An additional weakness is the healthy worker effect among this group of employees who have remained in employment until retirement age.[25] As this type of selection bias may influence the estimates of the associations between environmental exposures and disease risk, results from the study will need to be interpreted with caution.

Can I get hold of the data? Where can I find out more?

Currently, the study's data set is not freely available. However, we welcome ideas and proposals for potential collaborations. Any enquiries should be directed to T.W (wut@mails.tjmu.edu.cn) or H.Y. (yanghand@139.com) or F.B.H. (frank.hu@channing.harvard.edu).

Funding

Huazhong University of Science and Technology Foundation for Educational Development and Research; National Basic Research Program (Grant 2011CB503800); Natural Scientific Foundation of China (NSFC-81172751) and the DMC.

Acknowledgements

We thank all study participants and staff of the Health Examination Center of the Dongfeng Central Hospital and the Medical Insurance Center of DMC for their generous help. We also thank the interviewers from the retirement management office of DMC and from Tongji Medical College, HUST. The Tongji–Dongfeng Cohort Study is a collaboration among the Tongji Medical College, Huazhong University of Science and Technology, the DMC and the Harvard School of Public Health. F.W., J.Z., P.Y., X.L. and M.H. contributed equally to this work.

Conflict of interest: None declared.

KEY MESSAGES

- The response rate for the baseline survey of the Dongfeng–Tongji Cohort Study of Retired Workers was 87%. Most participants (98.2%) are Han Chinese.
- Mean age of the participants is 63.6 years and 35% have high school or higher education.
- Self-reported prevalence of chronic diseases at baseline is 38.1% for hypertension, 22.5% for high cholesterol, 11.8% for diabetes, 15.2% for CHD, 4.2% for stroke, 11.3% for chronic bronchitis and 3.3% for asthma.
- During their working lives, a high proportion of the participants were exposed to hazards at work: noise 12.5%, vibration 4.2%, dust 11.3%, high temperature 3.4% and organic toxins 15.3%.

References

[1] Wang L, Kong L, Wu F, Bai Y, Burton R. Preventing chronic diseases in China. *Lancet* 2005;**366:**1821–24.
[2] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;**3:** e442.
[3] Gu D, Reynolds K, Wu X *et al*. Prevalence of the metabolic syndrome and overweight among adults in China. *Lancet* 2005;**365:**1398–405.
[4] Jia WP, Wang C, Jiang S, Pan JM. Characteristics of obesity and its related disorders in China. *Biomed Environ Sci* 2010;**23:**4–11.
[5] Li JB, Wang X, Zhang JX *et al*. Metabolic syndrome: prevalence and risk factors in southern China. *J Int Med Res* 2010;**38:**1142–48.
[6] Li G, de Courten M, Jiao S, Wang Y. Prevalence and characteristics of the metabolic syndrome among adults in Beijing, China. *Asia Pac J Clin Nutr* 2010;**19:**98–102.
[7] Wang W, Kong J, Sun J *et al*. Epidemiological investigation of metabolic syndrome and analysis of relevant factors in north-eastern China. *J Int Med Res* 2010;**38:** 150–59.
[8] Hu FB, Liu Y, Willett WC. Preventing chronic diseases by promoting healthy diet and lifestyle: public policy implications for China. *Obes Rev* 2011;**12:**552–59.
[9] McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet* 1991;**337:**382–86.
[10] Chandalia M, Abate N, Garg A, Stray-Gundersen J, Grundy SM. Relationship between generalized and upper body obesity to insulin resistance in Asian Indian men. *J Clin Endocrinol Metab* 1999;**84:**2329–35.
[11] Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;**363:** 166–76.
[12] Cornelis MC, Agrawal A, Cole JW *et al*. The Gene-Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol* 2010;**34:**364–72.
[13] Zhang A, Park SK, Wright RO *et al*. HFE H63D polymorphism as a modifier of the effect of cumulative lead exposure on pulse pressure: the Normative Aging Study. *Environ Health Perspect* **118:**1261–66.

[14] Rybicki BA, Neslund-Dudas C, Nock NL *et al*. Prostate cancer risk from occupational exposure to polycyclic aromatic hydrocarbons interacting with the GSTP1 Ile105Val polymorphism. *Cancer Detect Prev* 2006;**30:** 412–22.
[15] Yuan BC, Su FM, Wu WT, Liu WS, Chiu KH. A predictive model of the association between gene polymorphism and the risk of noise-induced hearing loss caused by gunfire noise. *J Chin Med Assoc* 2012;**75:**36–39.
[16] Nebert DW, Roe AL, Vandale SE, Bingham E, Oakley GG. NAD(P)H:quinone oxidoreductase (NQO1) polymorphism, exposure to benzene, and predisposition to disease: a HuGE review. *Genet Med* 2002;**4:**62–70.
[17] Huang PC, Tsai EM, Li WF *et al*. Association between phthalate exposure and glutathione *S*-transferase M1 polymorphism in adenomyosis, leiomyoma and endometriosis. *Hum Reprod* 2010;**25:**986–94.
[18] Dawber TR, Kannel WB. An epidemiologic study of heart disease: the Framingham study. *Nutr Rev* 1958;**16:**1–4.
[19] Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. *Am J Nurs* 1978;**78:** 1039–40.
[20] Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *Br Med J* 1954; **1:**1451–55.
[21] Preis SR, Hwang SJ, Coady S *et al*. Trends in all-cause and cardiovascular disease mortality among women and men with and without diabetes mellitus in the Framingham Heart Study, 1950 to 2005. *Circulation* 2009;**119:**1728–35.
[22] Adams HP Jr, Bendixen BH, Kappelle LJ *et al*. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993;**24:**35–41.
[23] Anonymous. Nomenclature and criteria for diagnosis of ischemic heart disease. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. *Circulation* 1979;**59:**607–09.
[24] 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.
[25] Sterling TD, Weinkam JJ. The 'healthy worker effect' on morbidity rates. *J Occup Med* 1985;**27:**477–82.