BMJ Open

The National Cohort of Esophageal Cancer-Prospective Cohort Study of Esophageal Cancer and Precancerous Lesions based on High-Risk Population in China (NCEC-HRP): study protocol

Ru Chen,[1] Shanrui Ma,[2] Chentao Guan,[2] Guohui Song,[3] Qing Ma,[2] Shuanghua Xie,[2] Meng Wang,[2] Dantong Shao,[2] Xinqing Li,[2] Wenqiang Wei[1,2]

Check for updates

ABSTRACT

Introduction Oesophageal cancer (OC) is one of the most common cancers worldwide and about 50% of all new cases occurred in China. Population-based screening has been conducted in high-risk areas in China since 1970s, however, a few factors have limited the integration of the results from previous studies and the sharing of existing resources, such as the difference in screening methods and protocols, inconsistencies in questionnaires for risk factors investigation, lack of standards for sample collection and incomplete follow-up information.

Methods and analysis The National Cohort of Esophageal Cancer-Prospective Cohort Study of Esophageal Cancer and Precancerous Lesions based on High-Risk Population (NCEC-HRP) is a prospective cohort study of OC screening based on high-risk population in China supported by the National Key R&D Programme. Eight areas located at eastern, central and western China are selected as screening centres to represent three economical-geographical regions. All local residents aged 40–69 years in the selected areas are invited to take endoscopic examination and risk factors investigation unless they meet the exclusion criteria. The recruitment began on June 2017 and a total of 100 000 participants will be enrolled by December 2020 and all subjects will be followed for a long time. This study is designed as open-ended and has broad research aims. Summary statistics for baseline information will be reported after the completion of recruitment. We will develop a series of standards and guidelines for OC screening during the study. An open and shared research platform linked with epidemiological databases and biobank will be built up for further research.

Ethics and dissemination The study is approved by the Ethics Committee of Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (approval number 16-171/1250). The findings of the study will be disseminated through scientific peer-reviewed journals as well as the public via the study website (http://www.ncec-china.cn).

Trial registration number ChiCTR-EOC-17010553; Pre-results.

Strengths and limitations of this study

► The National Cohort of Esophageal Cancer-Prospective Cohort Study of Esophageal Cancer and Precancerous Lesions based on High-Risk Population is a prospective cohort of oesophageal cancer (OC) screening with large sample size, which is expected to enrol more than 100 000 subjects from eight regions in China.

► It is an open-ended study with broad research aims, comprehensive exposure data collection and long-term follow-up.

► A serious of standards on OC screening and relevant issues will be developed based on the high-quality evidence provided by this study.

► An open and shared research platform linked with epidemiological databases and biobank will be offered to researchers to conduct studies on OC and many other diseases.

► The study lacks of risk factors information from non-responders and the findings are limited to high-risk areas in China.

INTRODUCTION

Oesophageal cancer (OC) remains a significant source of morbidity and mortality worldwide. According to GLOBOCAN report, there are an estimated 572 034 new cases and 508 585 cancer deaths in 2018, with approximately a half of all new cases occurring in China.[1 2] OC has been the fifth most common cancer and the fourth most common cause of death in China.[3] The incidence rate and mortality of OC were 18.85/100 000 and 14.11/100 000, respectively, in China in 2014.[3] The two main types of oesophageal cancer are squamous cell carcinoma (OSCC) and adenocarcinoma, with OSCC accounting for over 90% of all cases of OC in.[4 5]

Although the exact cause of OC is unclear, it is considered as the result of a multiplicity

of demographic factors, diet and lifestyle, environmental and genetic factors. OC has a poor survival rate mainly result from the late stage at diagnosis. In China, its survival rate is less than 10% if diagnosed at an advanced stage but is as high as 85% if detected at an earlier stage.[6]

Since the 1970s, several screening programmes for OC have been conducted in the high-risk areas of China and achieved significant effects in the reduction of its incidence or mortality.[7–12] For example, Wei *et al* reported that the endoscopic screening and intervention significantly reduced mortality caused by OC during 10-year follow-up.[12] Two more high-quality randomised controlled trials are ongoing to evaluate the efficacy of endoscopic screening for OC in China.[13 14] However, a few factors have limited the integration of the results from previous studies and the sharing of existing resources, such as the difference in screening methods and protocols, inconsistencies in questionnaires for risk factors investigation, lack of standards for sample collection and incomplete follow-up information.

To this end, we started the National Cohort of Esophageal Cancer-Prospective Cohort Study of Esophageal Cancer and Precancerous Lesions based on High-Risk Population in China (NCEC-HRP) cohort, a prospective cohort study of OC and precancerous lesions based on high-risk population in China. As an important part of NCEC study, NCEC-HRP cohort focuses on populations in rural areas with high-incidence of OSCC. Learning from existing large cohorts such as the All of US Research Programme,[15] UK Biobank[16] and China Kadoorie Biobank,[17] NCEC-HRP is designed as a platform without a specific hypothesis to facilitate future research.

The major objectives of NCEC-HRP cohort are: (1) to establish a screening cohort based on high-risk population for OC; (2) to develop a series of standards and guidelines for OC screening, early diagnosis and treatment, risk factors investigation, sample collection and follow-up; (3) to build up a biobank with database on epidemiology, diagnosis, treatment and follow-up information and (4) to provide researchers with a platform for data sharing and promote communication and cooperation in OC research. Through standardised processes, NCEC-HRP could take advantage of resource integration, make rational use of samples and data, and provide scientific evidence for prevention and control of OC.

METHODS AND ANALYSIS
Study design and site selection
NCEC-HRP is a multi-centre prospective cohort study about OC screening. The study sites were carefully selected with the following criteria: (1) located at high-risk areas for OC; (2) representing different economical-geographical regions in China; (3) relative stability of the target population; (4) reliable local infrastructures including quality of existing cancer registry and death reporting systems, experienced doctors on OC screening, and availability of technology and equipment for sample collection and storage; (5) long-term local commitment to the project.

Figure 1 Location of selected sites and covering provinces for oesophageal cancer screening in China.

A total of eight sites are selected as the study centres (figure 1), including Feicheng of Shandong province, Yangzhong of Jiangsu province, Cixian of Hebei province, Linxian and Huaxian of Henan province, Yangcheng of Shanxi province, Yanting of Sichuan province and Chaoshan of Guangdong province. The incidence rates among selected areas ranged from 35.52/100 000 to 81.23/100 000.

Enrollment of participants

All men and women aged 40–69 years who are permanently residents in the selected study sites will be identified and invited to participate unless they met the following exclusion criteria: (1) history of cancer or mental disorder; (2) contraindications for endoscopic examinations and (3) inability to provide informed consent. The recruitment is conducted village by village. While propagandising the benefits of screening and early diagnosis, village doctors and local staff will also notify all target groups to go to designated hospitals for endoscopic examination. Those who are willing to do will be registered and scheduled for screening. If the hospital is far from the village, the vehicle will be arranged to pick up the participants to ensure the response rate. If the response rate of a village is too low (under 30%), a second mobilisation will be conducted to improve response rate. Participants' enrollment began on 1 June 2017 and a total of 100 000 subjects are expected to be included up to 31 December 2020.

Endoscopic examination and therapy

All endoscopic examinations and therapies are conducted by well-trained doctors at local hospitals. The screening procedure follows the recommendation of expert consensus on early OC screening and endoscopic diagnosis and treatment in China.[18] Briefly, screened participants are provided a standard upper gastrointestinal endoscopy with iodine staining. The entire oesophagus and stomach are visually examined and biopsies are taken from all focal lesions. Two pathologists independently read the biopsy slides without knowledge of the visual endoscopic findings. Discordances in the diagnosis are solved by discussion.

If early lesions are histologically diagnosed, participants will be recalled to the clinic and intervention methods appropriate to the lesions' severity would be used. For oesophageal severe dysplasia/carcinoma in situ or intramucosal carcinomas, endoscopic mucosal resection and/or endoscopic submucosal dissection treatments will be used as local therapies. For OCs, therapies include oesophagectomy, radical operation, radiotherapy and other conventional treatments.

Sample collection

Before endoscopy, blood will be collected into an EDTA containing vacutainer from each participant in the fasting state. A small portion is used for routine test for infectious diseases including HBV, HCV, HIV and syphilis, while the remaining (no less than 5 mL) will be dispensed into four

pipes after centrifugation, including two pipes of blood cell and two pipes of plasma. All samples will be stored at −80°C freezer for long-term preservation at each local site. Biopsy specimens are fixed in 10% buffered formalin and embedded in paraffin for storage. Pathological specimens and sections are preserved for patients with pathological examination. For each participate, saliva samples are also collected from the oral cavity by drooling and preserved in PreservCyt solution (Hologic, Bedford, Massachusetts, USA) at −70°C for use.

Follow-up and re-examination

We combine the active and passive follow-ups to accurately collect outcome information. For participants diagnosed with OC or precancerous lesions during the screening, annual interviews will be used to collect the outcome information through telephone, home visit or other contact methods. Meanwhile, a passive follow-up procedure will be carried out for all participants once a year. We will also collect additional data from local clinical settings, cancer registry system, death surveillance system, as well as medical insurance and claim databases to update the follow-up information. All participants will be followed for at least a decade. Figure 2 shows the enrollment, screening and follow-up procedure.

Re-examinations are required according to the diagnosis. For patients diagnosed with oesophageal mild dysplasia, a re-examination is required in 3 years and for those with oesophagus moderate dysplasia, an annual re-examination is required. For patients with severe dysplasia or in situ cancer, those who refuse treatment should be followed at least once a year.

Data collection

Although the cohort is based on OC screening population, we plan to use a uniform questionnaire based on the modified China Kadoorie Biobank (CKB) questionnaire to collect various exposure information,[17] because that CKB is one of the largest cohorts in the world for common chronic diseases, involving more than 0.5 million people in 10 regions of China.[19] Information collected through face-to-face interview will cover a broad range of variables (table 1), including demographic factors, indicators of socioeconomic status, smoking, alcohol and green tea consumption, diet, indoor air pollution, physical activity, reproductive history (women), sleep status, medical and family history. Some specific questionnaire were also developed according to the characteristics of OC, such as the history of digestive diseases, family history of cancer, drinking water, dietary habits (hot food, softness of food, eating speed), and oral hygiene. Blood pressure, heart rate, height, weight will also be measured.

In consideration of the mental health of screening subjects, we will conduct a pilot survey to investigate the impact of screening detected cancer on participants' psychosocial status. A serious of standardised instruments will be used to assess the patients' psychosocial status from multiple aspects, including General Anxiety Disorder-7
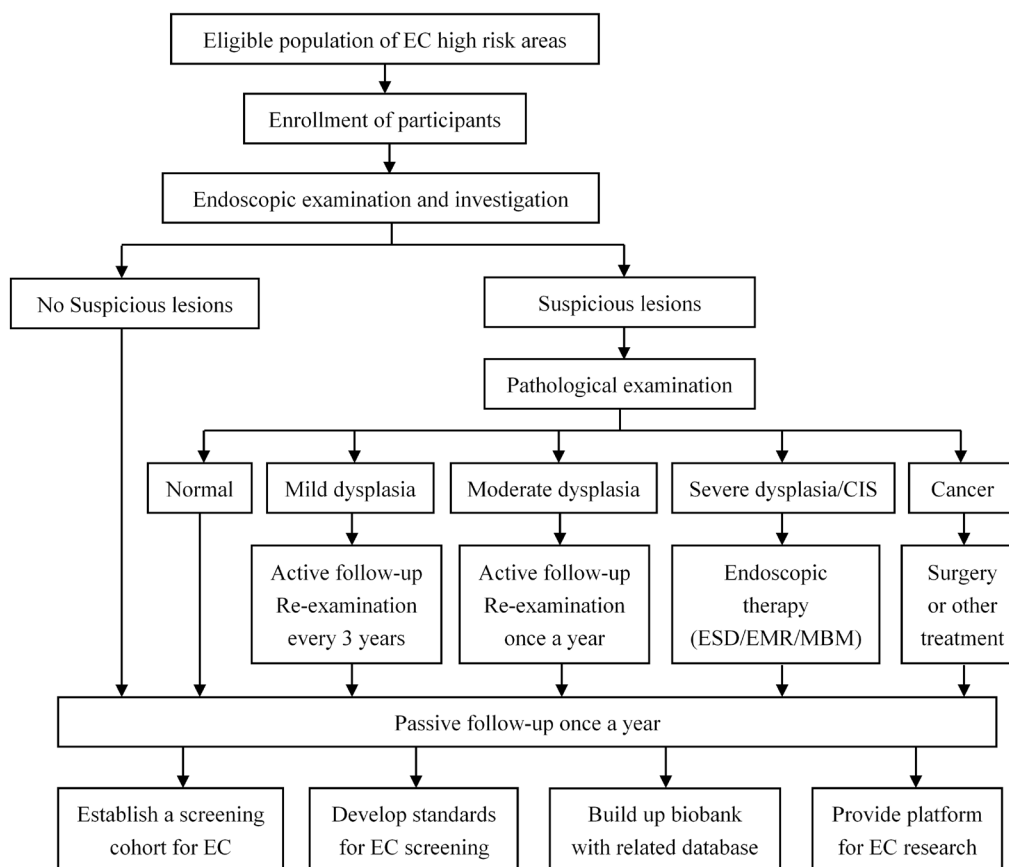
Figure 2 Enrollment, screening and follow-up procedure. O C , o esophageal cancer; CIS , carcinoma in situ ; ESD, endoscopic submucosal dissection; EMR, endoscopic mucosal resection; MBM, m ultiband mucosectomy.

for anxiety,[20] Patient Health Questionnaire-9 for depression,[21] The Connor-Davidson Resilience Scale for resilience,[22] Life Event Scale for life event stress,[23] Perceived Social Support Scale for social support[24] and Satisfaction With Life Scale for life satisfaction.[25]

Data management

All information collected is entered using a pad-based direct data entry app that was developed specifically for the project. The survey system has various built-in functions to avoid missing items and minimise logic errors during the interview. All data will be uploaded and stored into the data management platform at National Cancer Center. We aimed to design an open platform for professional research on OC and other health-related issues. The shared information contain data from questionnaires, diagnostic images from endoscopy and pathology, and biobank information.

We treated all data as protected health information and stored it securely as an encrypted and password protected database at the management centre. The raw data is backed up in both the pad-based survey system and the specific platform servers. The collection, shipping and receipt of data carriers were tracked by the management centre.

Outcome assessment

The primary objectives are to investigate the population distribution of OC and precancerous lesions in high-risk areas and to evaluate the effects of risk factors in a range of different circumstances. By storing blood, tissue and saliva samples, the study will facilitate reliable assessment on genetic factors and related mechanism research. With re-examination and long-term follow-up, dynamic changes in precancerous lesions can be observed and rare outcomes such as incident and death cases will also be accumulated, allowing us to further explore the transformation and progression of precancerous lesions. This will offer evidence for managing positive cases and optimising screening programmes.

Quality control

All investigators in the study are trained and assessed for consistency by investigating the same object with an experienced investigator. Ten percent of the participants will be randomly selected from the same village with repeat questionnaire and measures on selected items for quality control. During survey, the management centre will regular monitor the recruitment rate, the distribution of certain key variables and the sample collection through the system. The key quality assurance and quality control procedures in the study is summarised in online supplementary table 1.

Table 1 Summary of questionnaire data collected in the NCEC-HRP

Category	Example variables
Demographic information	Gender, race, birthday, address, contact number
Socioeconomic information	Marriage, education, occupation, number of house member, house income, insurance type
Behavioural factors	Alcohol, smoking, green tea, drinking water, consumption of certain food (fresh vegetables, meat/poultry, fish/sea food, egg, soybean, dairy products, beverages, pickled food, mouldy food, spicy food), dietary habits (hot food, softness of food, eating speed), nutritional supplement, physical activity
General health related information	Self-rated health status (self-reported health status, oral hygiene, history of trauma, history of cancer and related therapy, history of digestive disease, *Helicobacter pylori* infection status, history of chronic disease), current medication (non-steroidal anti-inflammatory drugs, steroidal anti-inflammatory drugs, acid inhibitor, antibiotics), exposure to indoor air pollution from cooking/heating fuel, exposure to passive smoking, sleep situation
Family history	Parental age/or age of death, number of siblings, number of children, family history of cancer, family history of chronic disease
Reproductive history (for women)	Age of first menstrual period, menopause status, history of contraceptive pills use, history of hysterectomy and of ovary/breast surgery

Sample size calculation

According to previous studies, the proportion of precancerous lesions in target screening population in high risk areas is about 20%–25%. Within the prospective cohort that expected to enrol 100 000 participants, there will be more than 20 000 cases with precancerous lesions. Assuming the detection rate of precancerous lesions is 20%, to achieve a precision of 2% with an α of 0.05, it would need a sample size of 3252 to achieve a power of 80%. Therefore, we will has adequate power for all type of precancerous lesions. In the exploring of risk factors for precancerous lesions, for a factor of 10% exposure level, we could also detect a quite small effect (OR of 1.1) with 80% power at the 5% level of significance.

Data analysis plan

First, we will present summary statistics for baseline variables including demographic, socioeconomic and behavioural characteristics, general health-related information and family history. Senconde, the detection rate and the distribution of OC and precancerous lesions among different population characteristics (eg, age group, gender, site, screening year etc) will be reported.

Third, parametric and nonparametric tests such as t-test, χ^2-test, Fisher's exact test and Wilcoxon rank sum test will be applied for univariate analyses to identify risk factors associated with interested outcomes. Fourth, multivariable regression analyses, including linear, logistic, Cox proportional hazard and Poisson models will be selected as appropriate to assess the association between risk factors and outcomes with adjustment for potential confounders. The tests will be two sided and $p < 0.05$ will be considered statistically significant. Data analysis will be undertaken using STATA V.14.0.

Ethics and dissemination

The transmission, storage and analysis of health-related personal data and the storage of biological samples within this project will strictly follow the legal requirements for data protection under the supervision of the ethics committee. Data protection and confidentiality will be guaranteed.

The findings of the study will be disseminated through scientific peer-reviewed journals as well as the public via the NCEC study website (http://www.ncec-china.cn). The NCEC study group is committed to making the cohort data available to the scientific researchers worldwide to produce advance knowledge about the causes, prevention and treatment of OC and other diseases. Researchers with related interests are invited to contact us for further discussion. More information on date sharing and application will be published online in the future.

Patient and public involvement

The protocol of this study was discussed and developed by a multidisciplinary team of experts including epidemiologists, clinicians, statisticians and computer engineers as well as field investigators, but no patients or public were involved in the design phase.

DISCUSSION

Strengths of the study include the prospectively collected exposure data, an exceptionally large sample size and the opportunity to follow all participants through various resources. Based on the high quality evidence provide by the study, we would also develop a serious of standards and guidelines on OC screening, early diagnosis and treatment, risk factors investigation, sample collection and follow-up. Those achievements might be implemented to other high risk areas, especially in developing countries in Central and East Asia and Eastern Africa. In addition, the biobank with blood, tissue and saliva will lay the foundation for further exploration on the genetic risk factors and relevant mechanism of OC and other diseases. Finally, as an open-ended prospective study, the exposure measurements are not only designed for OC, but also for many other chronic diseases. Through linking epidemiological databases tothe biobank, NCEC-HRP will also provide an open and shared research platform for researchers worldwide to conduct studies on OC and many other diseases.

A major limitation in this study is the lack of risk factors information from non-attenders. However, a questionnaire and sampling method for non-attenders is already under design. We also try to obtain the baseline and outcome information of non-participants by referring to various date sources to improve the accuracy of the research results from multiple aspects. Second, our cohort are limited to high-incidence areas, which might affect its extrapolation in low-incidence areas.

The pre-existing cohorts and ongoing trials bring new insights into OC, however, the lack of standardisation and sharing limit the popularisation and application of results. A large cohort is urgently needed which should own uniform standards on OC screening, early diagnosis and treatment, risk factors investigation, sample collection and follow-up. Our cohort is aimed to establish databases and biobank in order to offer sufficient samples and complete data for subsequent multi-omics researches. With large sample size and long-term follow-up, we could achieve a more comprehensive understanding of the epidemiology of OC and precancerous lesions, including the incidence and mortality, risk factors, progress and survival and so on. NCEC-HRP is part of the NCEC study and, together with other four cohorts including standardised diagnosis and treatment of clinical cohort, minimally invasive treatment of early stage and precancerous lesion cohort, genetic lineage cohort and prospective cohort based on urban community, will construct a comprehensive research platform for OC and precancerous lesions.

Author affiliations
[1]Cancer Registry Office, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[2]Department of Cancer Epidemiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[3]Cixian Institute for Cancer Prevention and Control, Handan, China

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
2. International Agency for Research on Cancer. In: BW S, CP W, eds. *World Cancer Report 2014*. Lyon: World Health Organization Press, 2014.
3. Chen W, Sun K, Zheng R, et al. Cancer incidence and mortality in China, 2014. *Chin J Cancer Res* 2018;30:1–12.
4. Abnet CC, Arnold M, Wei WQ. Epidemiology of esophageal squamous cell carcinoma. *Gastroenterology* 2018;154:360–73.
5. Arnold M, Soerjomataram I, Ferlay J, et al. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* 2015;64:381–7.
6. Wang GQ, Jiao GG, Chang FB, et al. Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening. *Ann Thorac Surg* 2004;77:1740–4.
7. Dawsey SM, Fleischer DE, Wang GQ, et al. Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China. *Cancer* 1998;83:220–31.
8. Dawsey SM, Wang GQ, Weinstein WM, et al. Squamous dysplasia and early esophageal cancer in the Linxian region of China: distinctive endoscopic lesions. *Gastroenterology* 1993;105:1333–40.
9. Qin DX, Wang GQ, Zuo JH, et al. Screening of esophageal and gastric cancer by occult blood bead detector. *Cancer* 1993;71:216–8.
10. Roth MJ, Liu SF, Dawsey SM, et al. Cytologic detection of esophageal squamous cell carcinoma and precursor lesions using balloon and sponge samplers in asymptomatic adults in Linxian, China. *Cancer* 1997;80:2047–59.
11. Shen O, Liu SF, Dawsey SM, et al. Cytologic screening for esophageal cancer: results from 12,877 subjects from a high-risk population in China. *Int J Cancer* 1993;54:185–8.
12. Wei WQ, Chen ZF, He YT, et al. Long-term follow-up of a community assignment, one-time endoscopic screening study of esophageal cancer in China. *J Clin Oncol* 2015;33:1951–7.
13. He Z, Liu Z, Liu M, et al. Efficacy of endoscopic screening for esophageal cancer in China (ESECC): design and preliminary results of a population-based randomised controlled trial. *Gut* 2019;68.
14. Chen W, Zeng H, Chen R, et al. Evaluating efficacy of screening for upper gastrointestinal cancer in China: a study protocol for a randomized controlled trial. *Chin J Cancer Res* 2017;29:294–302.
15. Precision Medicine Initiative (PMI) Working Group. The precision medicine initiative cohort program-building a research foundation for 21st century medicine. 2015 file:///C:/Users/Chenru/Desktop/pmi-working-group-report-20150917-2.pdf.
16. Hewitt J, Walters M, Padmanabhan S, et al. Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. *BMJ Open* 2016;6:e009161.
17. 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.
18. Chinese Medical Association Chinese Society of Digestive Endoscopy. Chinese expert consensus on screening and endoscopic management of early Esophageal Cancer (Beijing, 2014). *Chinese Journal of Gastroenterology* 2015;20:220–40.
19. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;40:1652–66.
20. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;166:1092–7.
21. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;282:1737–44.
22. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depress Anxiety* 2003;18:76–82.
23. Mooy JM, de Vries H, Grootenhuis PA, et al. Major stressful life events in relation to prevalence of undetected type 2 diabetes: the Hoorn Study. *Diabetes Care* 2000;23:197–201.
24. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24:385–96.
25. Diener E, Emmons RA, Larsen RJ, et al. The satisfaction with life scale. *J Pers Assess* 1985;49:71–5.