Hypertension Research (2010) 33, 722–726

www.nature.com/hr

ORIGINAL ARTICLE

Evaluating the risk of hypertension using an artificial neural network method in rural residents over the age of 35 years in a Chinese area

Shuqiong Huang[1], Yihua Xu[1], Li Yue, Sheng Wei, Li Liu, Xiumin Gan, Shuihong Zhou and Shaofa Nie

Hypertension (HTN) has been proven to be associated with an increased risk of cardiovascular diseases. The purpose of the study was to examine risk factors for HTN and to develop a prediction model to estimate HTN risk for rural residents over the age of 35 years. This study was based on a cross-sectional survey of 3054 rural community residents (*N*=3054). Participants were divided into two groups: a training set (*N*1=2438) and a validation set (*N*2=616). The differences between the training set and validation set were not statistically significant. The predictors of HTN risk were identified from the training set using logistic regression analysis. Some risk factors were significantly associated with HTN, such as a high educational level (EL) (odds ratio (OR)=0.744), a predominantly sedentary job (OR=1.090), a positive family history of HTN (OR=1.614), being overweight (OR=1.525), dysarteriotony (OR=1.101), alcohol intake (OR=0.760), a salty diet (OR=1.146), more vegetable and fruit intake (OR=0.882), meat consumption (OR=0.787) and regular physical exercise (OR=0.866). We established the predictive models using logistic regression model (LRM) and artificial neural network (ANN). The accuracy of the models was compared by receiver operating characteristic (ROC) when the models were applied to the validation set. The ANN model (area under the curve (AUC)=0.900 ± 0.014) proved better than the LRM (AUC=0.732 ± 0.026) in terms of evaluating the HTN risk because it had a larger area under the ROC curve.
Hypertension Research (2010) **33,** 722–726; doi:10.1038/hr.2010.73; published online 27 May 2010

Keywords: prediction model; risk factors; rural

INTRODUCTION

Hypertension (HTN) is one of the leading causes of cardiovascular and cerebrovascular diseases in the aging population[1] and is related to the changing of lifestyle in China. Cerebrovascular disease became the second leading cause of death for urban residents and the top cause in rural areas of China in 1998.[2] The prevalence of HTN has increased in China; 27.2% of Chinese adults aged 35 to 74 years have HTN.[3]

Hypertension is a great threat to people's physical and mental health, but it can be prevented and controlled.[4] Estimating an individual's risk across a range of presumed risk factors is fundamental to prevent HTN.[5] Due to its complex, multifactorial nature, the prevention of HTN must refer to multiple risk factors.[6,7] Evidence suggests that the risk for progression to HTN depends on risk factors, such as baseline blood pressure, age, body mass index (BMI),[8,9] family history of HTN (FH), occupation, physical exercise (PE), smoking, alcohol intake (AI) and dietary pattern.[10] High-risk individuals will probably derive the maximal benefit from non-pharmacologic (lifestyle-related) interventions aimed at preventing HTN.[11–14] Thus, we believe that knowledge of the risk for HTN will aid in patient education and counseling, and it will assist in clinical decision making and design of future interventional studies.

Despite the rapid transition in China, the development is not equal between urban and rural areas.[15] Inequalities between the regions have manifested, and rural adults may receive fewer routine checkups or preventive services, including screening for HTN.[16] As a result, evaluating the HTN risk for rural residents is critically important.

The purpose of this study was twofold: (1) we examined the risk factors for HTN based on a cross-sectional study by logistic regression analysis and (2) we attempted to develop a prediction model that can be used in primary health care to estimate the probability that an individual will develop HTN for rural residents aged over 35 years in a Chinese area.

METHODS

Study population

The study was based on a cross-sectional survey. All participants filled out a questionnaire including demographic, lifestyle and dietary data. Some measured traits, such as blood pressure, height, and weight, were tested by trained health

Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, PR China
[1]These authors contributed equally to this work.
Correspondence: Professor S Nie, Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Hangkong Road 13, Wuhan 430030, PR China.
E-mail: tjnsf2007@163.com
Received 10 September 2009; revised 19 February 2010; accepted 7 March 2010; published online 27 May 2010

staff at the same time. The study was designed and carried out at Tongji Medical College in a rural area of Yichang Three Gorge Dam in the Hubei Province.

The participants for the study were chosen from the rural residents in a three-step process. First, individuals under 35 years old were excluded. Second, those who could not cooperate with the investigation because of long-term outing or illness were excluded from the remaining candidates. Third, we excluded essential HTN patients treated with antihypertensive medication at the time of investigation and those suffering from known coronary heart disease, diabetes mellitus or chronic renal disease. Ultimately, a total of 3087 residents who met the criteria were originally enrolled in the study. Informed consent was obtained from all participants. The research was approved by the Tongji Medical College ethics committee.

Data collection

Data collection included the questionnaire and physical measurements (height, weight and blood pressure). The questionnaire consisted of sociodemographic characteristics and dietary and lifestyle behaviors. Data pertaining to blood pressure, height and weight were tested for all participants by trained health-care staff at the same time. Data were obtained from a total of 3054 (N=3054) participants. The response rate was 98.9%. The measurements were taken by trained health-care staff. The questionnaire was designed by the scientists on our research team. Before the formal investigation, the questionnaire was validated and revised repeatedly according to the analysis of the results.

Assessment of risk factors

Sociodemographic status. The sociodemographic status of participants was classified into different categories on the basis of occupation and educational level (EL). Education level was classified into five categories: illiterate, elementary level, secondary school, high school and college. The latter two categories were grouped as high EL. Occupations included farmers, laborers, professionals and employers/managers. The last occupation was defined as predominantly sedentary work (PSW). Positive FH was defined as documented maternal and paternal HTN at or before the baseline examination.

Anthropometric measurements. Blood pressure difference is the difference between systolic blood pressure (SBP) and diastolic blood pressure (DBP). The normal range of blood pressure difference is from 30 to 40 mm Hg. Dysarteriotony is defined as a pressure difference higher than 60 mm Hg or lower than 20 mm Hg. Three blood pressure readings were taken in a quiet sitting position using a mercury sphygmomanometer with a 2-mm Hg scale according to the standard procedure.[17] The mean of the three readings was used for data analysis.

Body mass index (BMI), used as an index of relative weight, was calculated as body weight (in kilograms) divided by the square of height (in meters). Height and weight were measured following a standard methodology.[18] The data on height and weight were based on the mean of two measurements. The Chinese BMI categories established in 2002 were used as the criteria for defining overweight (BMI was 24 or higher but less than 28) and obese (BMI was 28 or higher) patients.[19] Respondents were considered overweight if their BMI was 24 or higher.

Dietary and lifestyle behaviors. Dietary patterns consisted of salt intake, vegetable and fruit intake (VFI), fat intake and other data. A salty diet (SD) is defined as consuming an average of more than 6 g per day. Fat intake is defined as consuming an average of more than 25 g per day.

Many studies on smoking exist, with various definitions of smoking status.[20–24] In this study, individuals who currently smoked and had smoked at least 100 cigarettes during their lifetime were defined as current smokers, which is the same definition used by the community intervention trial for smoking cessation (COMMIT) in the United States.[21] In our survey, self-reported smoking was measured by the following question: 'Do you smoke any cigarettes, self-made cigarettes or cigars now?' and 'Have you smoked at least 100 cigarettes during your lifetime?'

Alcohol consumption data were also collected from the questionnaire using the following question: 'Considering all types of alcoholic beverages, how many times during the past 30 days did you drink?' For men, moderate drinking is defined as consuming an average of two drinks or less per day and heavy drinking is defined as consuming an average of more than two drinks per day. For women, moderate drinking is defined as consuming an average of one drink or less per day and heavy drinking is defined as consuming an average of more than one drink per day.[25,26] Therefore, in our survey, self-reported drinking was measured by the following question: 'Do you consume an average of more than 1 drink per day?'

Physical activity was also addressed by the questionnaire. On the basis of the definition of physical activity in the Behavioral Risk Factor Surveillance System (BRFSS) from the United States,[27] physical activity was defined as follows: physical work, walking, riding bicycle, jogging, dancing or Qigong for 30 or more minutes per day for 6 days or more per week. The respondents were asked how many days per week they engaged in at least 30 min of physical activity, including physical work, walking, riding bicycle, jogging, dancing or Qigong, according to the following categories: 6–7 days/week, 3–5 days/week, 1–3 days/week or never. The first two categories were grouped together as regular PE.

Definition of HTN

Subjects were considered normal if their average SBP and DBP were less than 140 mm Hg and less than 90 mm Hg, respectively, at the time of examination and if they had no previous diagnosis of HTN and were not treated with antihypertensive medication at the time of examination. Subjects were considered hypertensive if their average SBP or DBP were greater than 139 mm Hg or greater than 89 mm Hg, respectively, at the time of examination.

Statistical analysis

In this study, 3054 participants, according to gender and age group (from 35 years old with 5-year intervals; those 65 years old and above were merged as one group) indicators, were randomly divided into two parts as a training set ($N1$=2438) and validation set ($N2$=616) in a 4:1 pair matched ratio. The proportion of normal blood pressure and high blood pressure in the training set and validation set was the same as that in the population data, and there was no statistically significant difference by χ^2-test for gender and age between the two data sets. The training set was used to select variables and to establish a model. Then the validation set was used to test and evaluate the model.

Analysis was undertaken in three stages. In the first stage, a set of predictors contributing to the prediction of HTN risk was identified by logistic regression analysis based on the training set. Continuous variables were divided into categories to facilitate risk estimation.

In the second stage, prediction models for a binary outcome variable specified the probability of having HTN using a logistic regression model (LRM) and an artificial neural network (ANN) model. The models were performed with the probability of having HTN as the dependent variable and the risk factors as the independent variable. A multivariate LRM for a binary outcome variable specified the logit (logarithm of the odds) of the probability of having HTN to be a function of a set of predictors. An ANN, usually called a 'neural network', is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks.[28,29] Neural networks are nonlinear statistical data modeling tools that can be used to model complex relationships between inputs and outputs or to find patterns in data. There are several types of neural networks, including the feed forward neural network, the radial basis function (RBF) network and the Kohonen self-organizing network. The feedforward neural networks, including back-propagated delta rule networks (BP) and other networks, were the first and arguably simplest types of ANNs devised. Although there are different neural networks, BP networks are the most popular choice due to their relatively simplicity and stability.[30,31] In general, a neural network consists of three layers: an input layer to accept information, a hidden layer to process information and an output layer to calculate responses. In our study, the ANN model (BP network) used the same inputs as the LRM.

In the third stage, we assessed performance of the risk evaluation model by using receiver operating characteristic (ROC) curve analysis.[32] The area under the curve is a measure of test accuracy. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test (or the method). On the other hand, the closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test. The larger the area under the curve, the greater the accuracy of the model.

Logistic regression analysis and ROC curve analysis were performed using SPSS13.0 (SPSS, Chicago, IL, USA). The ANN model was performed with MATLAB 7.1. All reported *P*-values were two-sided, and *P*-values less than 0.05 were considered to be statistically significant.

Determination of the cut-off point

Because the evaluations were probabilities, we had to first adopt an arbitrary cut-off point to classify the predicted probabilities. The probability of having HTN was denoted as '*P*' in the model. The cut-off point was used to classify whether the evaluated probabilities belonged to the positive prediction or to the negative prediction. We were aware of two problems associated with the approach. First, different cut-off points could have resulted in different classification rates. Second, the classification rate derived from one study could have been different from other results reported by different studies, even with the same set of predictors and the same participants, unless there was a consensus on the adopted cut-off points. In this study, we arbitrarily adopted a cut-off point at 0.5 for this purpose; that is, participants were identified to be at high risk for HTN when $P \geqslant 0.5$ but not when $P < 0.5$.

RESULTS

Description of the participants

The age of the participants was over 35 years, with a mean age of 51.43 ± 12.22 years in the training set and 51.39 ± 11.82 years in the validation set. The training set had 1133 males and 1305 females (53.5%), and there were 311 males and 305 females (49.5%) in the validation set. The rural participants mainly received primary or middle school education. The occupation of the participants was predominantly farmers. The mean blood pressure of the participants was 126.5 ± 20.2 mm Hg for SBP and 77.2 ± 11.38 mm Hg for DBP in the training set and 126.6 ± 20.48 mm Hg for SBP and 76.8 ± 10.99 mm Hg for DBP in the validation set. A total of 655 individuals of the 2438 participants (26.9%) in the training set were hypertensive. A total of 168 individuals of the 616 participants (27.3%) in the validation set were hypertensive. The differences between the training set and the validation set were not substantial and were not statistically significant in terms of age, gender, EL, occupation or blood pressure. Participants in each set had similar baseline characteristics. Table 1 shows the characteristics of the study participants.

Predictors of HTN risk

Table 2 shows the predictors of HTN risk identified from logistic regression analysis based on the training set ($N1=2438$). With the construction of a multivariable model (logistic regression analysis), we found some factors to be significantly associated with HTN risk. In our study, we found a significant positive relationship between the risk factors and HTN, including being overweight (BMI ≥ 24 kg m^{-2}) (odds ratio (OR)=1.525), having a positive FH (OR=1.614), consuming a SD (OR=1.146), performing predominantly sedentary work (PSW, OR=1.090) and having dysarteriotony (BP, OR=1.101). There was also an inverse relationship between the five protective factors, including high EL (OR=0.744), AI (OR=0.760), more VFI (OR=0.882), meat consumption (AII, OR=0.787) and regular PE (OR=0.866), and the prevalence of HTN.

Prediction models

The resulting logit probability of having HTN was described by the following linear LRM (we denoted it as LRM):

$$
\begin{aligned}
&- 5.228+0.087X_{PSW}+0.479X_{FH} - 0.296X_{EL} \\
&- 0.275X_{AI} - 0.125X_{VFI}+0.136X_{SD} - 0.240X_{AII} \\
&- 0.143X_{PE}+0.422X_{BMI}+0.096X_{BP}
\end{aligned}
$$

Table 1 Baseline characteristics of the participants

Characteristic	Participants (N=3 054) Training set (N1=2 438)	Validation set (N2=616)
Mean age (s.d.), years	51.43 (12.22)	51.39 (11.82)
Women (%)	1305 (53.5)	305 (49.5)
Mean blood pressure (s.d.), mm Hg		
Systolic	126.5 (20.2)	126.6 (20.48)
Diastolic	77.2 (11.38)	76.8 (10.99)
Educational level (%)		
No education	369 (15.1)	96 (15.6)
Primary school	1050 (43.1)	271 (44)
Middle school	826 (33.9)	204 (33.1)
High school	109 (6.9)	45 (7.3)
College	9 (0.4)	0 (0)
Occupation (%)		
Farmers	2039 (83.6)	503 (81.7)
Laborers	278 (11)	79 (12.9)
Employers/managers	109 (4.4)	27 (4.4)
Current smoker (%)	895 (36.7)	233 (37.8)
Current drinker (%)	676 (27.7)	193 (31.3)
Physical activity (%)	485 (19.9)	
Family history of hypertension (%)	1935 (79.4)	503 (81.7)
Mean body mass index (s.d.), kg m^{-2}	22.1 (2.95)	22.3 (3.11)
Hypertension (%)	655 (26.9)	168 (27.7)

Table 2 Predictors of hypertension risk for an individual

Variable	β	s.e.	P-value[a]	OR (95% CI)
Predominantly sedentary work, X_{PSW}	0.087	0.040	0.030	1.090 (1.008–1.179)
Positive family history, X_{FH}	0.479	0.170	0.005	1.614 (1.157–2.252)
High-educational level, X_{EL}	−0.296	0.068	0.000	0.744 (0.650–0.850)
Alcohol intake, X_{AI}	−0.275	0.123	0.025	0.760 (0.597–0.967)
More vegetable and fruit intake, X_{VFI}	−0.125	0.060	0.036	0.882 (0.784–0.992)
Salty diet, X_{SD}	0.136	0.045	0.002	1.146 (1.051–1.250)
Animal insides intake, X_{AII}	−0.240	0.063	0.000	0.787 (0.696–0.889)
Regular physical exercise, X_{PE}	−0.143	0.083	0.086	0.866 (0.736–1.020)
Overweight, X_{BMI}	0.422	0.131	0.001	1.525 (1.181–1.971)
Dysarteriotony, X_{BP}	0.096	0.005	0.000	1.101 (1.091–1.112)

Abbreviations: AI, alcohol intake; BMI, body mass index; BP, back-propagated delta rule networks; CI, confidence interval; EL, educational level; FH, family history of hypertension; OR, odds ratio; PE, physical exercise; PSW, predominantly sedentary work; SD, salty diet; VFI, vegetable and fruit intake.
[a]*P*-value for test of the relation between risk factors and hypertension.

We built the ANN model on the basis of the predictors of HTN risk resulting from the logistic regression analysis. The predictors were used as the model input: BMI, FH, SD, occupation, blood pressure difference, EL, AI, VFI, meat consumption and PE. The binary variable of whether an individual was suffering from HTN was employed as the output variable. Only two input variables were continuous variables—BMI and blood pressure difference—whereas the others were categorical variables. The analysis structure of the BP neural network included three layers (Figure 1), namely, the input

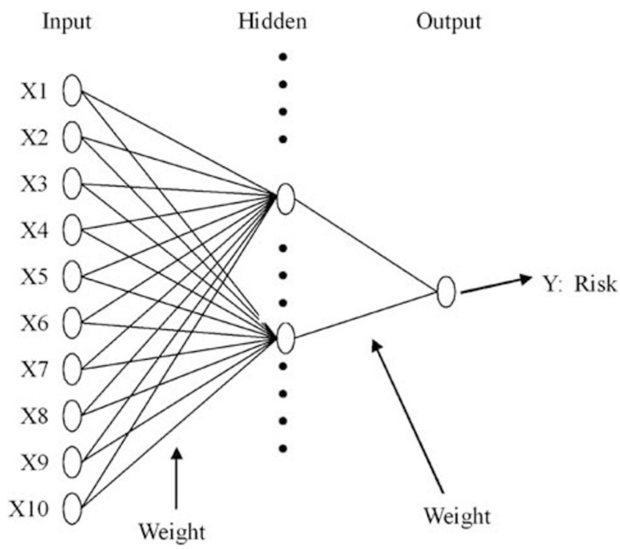

Figure 1 Three layers feedforward network for evaluating hypertension risk of an individual. X1, occupation; X2, family history; X3, educational level; X4, alcohol intake; X5, vegetable and fruit intake; X6, salty diet; X7, animal insides intake; X8, physical exercise; X9, BMI; X10, blood pressure difference.

layer with 10 neurons, the 21 hidden layer neurons and the output layer neurons, corresponding to the forecast variable (that is the probability of having HTN).

Discriminatory ability of models

The two models could successfully distinguish the individual risk of having HTN. We compared the individual risk output from the two models with the actual status using the validation set ($N2=616$) (Table 3). The ANN model detected 160 HTN patients from the actual hypertensive patients, whereas the LRM only detected 85. The sensitivity of ANN (0.952) was higher than that of LRM (0.506).

Figure 2 summarizes the ROC curve areas obtained from the LRM and the neural network model. The areas under the ROC curves were different between the two models. The areas under the ROC curves (area under the curve) were 0.900 ± 0.014 for ANN and 0.732 ± 0.026 for LRM. The area under curve of the ANN model was significantly higher than that of the LRM. Our results showed that the neural network model had a better predictive performance than the LRM. To provide more insight into the comparison, we also examined other indicators in the study, including sensitivity and specificity. We found that sensitivity, specificity and Yourdon's index at the adopted cut-off point for the LRM model (and ANN model) were 0.51 (0.95), 0.96 (0.85) and 0.46 (0.80), respectively. Thus, the ANN was selected as our 'better' model because it had the larger area under the ROC curve and the higher Yourdon index.

Table 3 Comparison of the individual risk results from the two models and the actual status using validation set

	LRM		*ANN*		
Actual status	*HT (%)*	*NT*	*HT (%)*	*NT*	*Total*
Hypertension (HT)	85 (50.6)	83	160 (95.2)	8	168
Normotensive (NT)	19	429	68	380	448
Total	104	512	228	388	616

Abbreviations: ANN, artificial neural network; LRM, logistic regression model.

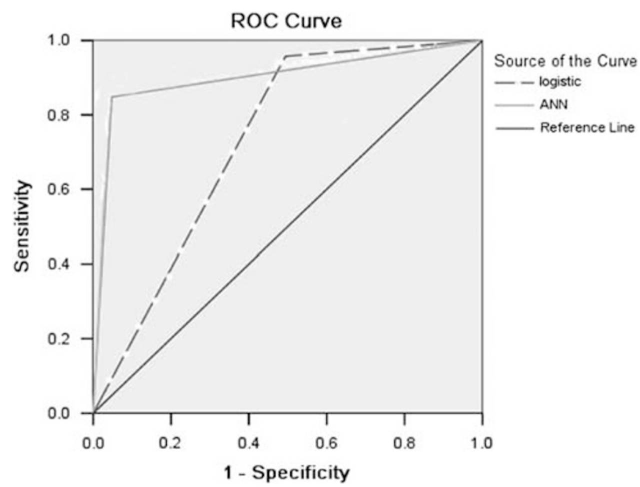

Figure 2 The ROC curves are obtained from the logistic regression model (LRM) and the artificial neural network model (ANN). *X* axis: sensitivity, *Y* axis: 1—specificity. The reference line is shown in solid line (the right). The ROC curve of ANN model is simply shown in dashed line (the left), and the ROC curve of LRM in dotted line (the middle). The area under the dashed line indicates larger than that under the dotted line. ANN has the larger area under the ROC curve than LRM.

DISCUSSION

There were two fundamental questions that we sought to answer in this study: (1) whether there is a relationship between the individual-related factors and HTN and (2) whether the neural network approach could provide a more accurate evaluation to identify persons who are at the highest risk for HTN.

To answer our first question, we examined the distributions of the predictors of HTN risk from our analysis. Only occasionally were some factors in the model, or the different interpretations of the same variables, quite different from other research. For example, the categorization of occupation used in this study was composed of farmers, non-farmers and employers/managers. One study defined occupation as a dichotomous variable: heavy work activity (=1) and light-to-moderate work activity (=0), based on the physical nature of the work.[33] In addition, AI was a protective factor for HTN in our study, which is in conflict with other research findings. Some researcher advise lifestyle modifications, such as cessation of drinking, healthy diet and moderate physical activity, to prevent or delay the progression to HTN in rural areas of China.[34] This discrepancy was probably related to the reporting bias of the participants, and we did not distinguish the amount of the AI in this survey.

To answer our second question, we considered accuracy performance by examining discrimination. Our results showed that the distributions of individual HTN risk obtained from our models were close to the distributions of actual status in the participants. On the results obtained in the study, there was some evidence to suggest that neural networks produce more accuracy. The conclusion drawn from the comparative study of the two models was certainly credible.

As in all modeling approaches, there is always a tendency that we might over fit the data. That is, the fitted model may have been too specific for the training data set and will not generalize well to other independent data sets. The ANN network has already been used in oncology, diabetes and other diseases[35–39] and has shown its unique advantages. In this study, the ANN method used a hidden layer with 21 neurons, which not only reduced the amount of calculation but also prevented over-fitting. We confirmed that the ANN model had a stronger performance in predicting HTN risk for a rural individual aged over 35 years.

However, there were some limitations in our study. The samples were limited geographically and ethnically,[40] consisting of a rural community of individuals aged over 35 years. This cross-sectional design prohibits any inferences of a causal association between the risk factors and HTN. We measured the variables, such as BMI, AI, VFI, meat consumption and PE, on only a single occasion and did not assess changes in these variables. We also considered some lifestyle variables to be dichotomous variables rather than continuous variables. We chose this approach to keep the analysis simple. We also measured blood pressure on a single occasion, which may not be as accurate as several measurements. Finally, our prediction model was based on a sample design. Despite these limitations, we developed an accurate risk predictive model that estimates the combined impact of an individual's related factors on HTN.

In conclusion, we examined risk factors for HTN based on a cross-sectional study and developed a simple risk evaluation model using a neural network-based method that estimates the HTN risk for an individual. We believe that this result may facilitate targeting of prevention interventions, which is a key objective for individuals at risk for HTN. The prediction model has some flaws because of the limitations of the study areas and restrictions to selected variables. Studies are needed to further validate whether our findings are applicable in other samples.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported in part by the China '863' high-tech research and development project (Grant number 2006AA02Z347). We also thank all of the participants who took part for their support and all of the coordinators for their help.

1 Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; **360**: 1903–1913.
2 China Hypertension League, Revising Committee for The Guidelines. *Guidelines for the management of hypertension of China*. Health Ministry of People's Republic of China: Beijing, 2005.
3 Gu DF, Reynolds K, Wu XG, Chen J, Duan XF, Muntner P, Huang GY, Reynolds RF, Su SY, Whelton PK, He J. Prevalence, awareness, treatment, and control of hypertension in China. *Hypertension* 2002; **40**: 920–927.
4 Kaplan NM. Anxiety-induced hyperventilation—A common cause of symptoms in patients with hypertension. *Arch Intern Med* 1997; **157**: 945–948.
5 Arima H, Yonemoto K, Doi Y, Ninomiya T, Hata J, Tanizaki Y, Fukuhara M, Matsumura K, Iida M, Kiyohara Y. Development and validation of a cardiovascular risk prediction model for Japanese: the Hisayama study. *Hypertens Res* 2009; **32**: 1119–1122.
6 Gueyffier F, Boissel JP, Pocock S, Boutitie F, Coope J, Cutler J, Ekbom T, Fagard R, Friedman L, Kerlikowske K, Perry M, Prineas R, Schron E. Identification of risk factors in hypertensive patients: contribution of randomized controlled trials through an individual patient database. *Circulation* 1999; **100**: e88–e94.
7 Chiong JR. Controlling hypertension from a public health perspective. *Int J Cardiol* 2008; **127**: 151–156.
8 Vasan RS, Larson MG, Leip EP, Kannel WB, Levy D. Assessment of frequency of progression to hypertension in non-hypertensive participants in the Framingham Heart Study: a cohort study. *Lancet* 2001; **358**: 1682–1686.
9 Pang WY, Sun ZQ, Zheng LQ, Li J, Zhang XG, Liu SS, Xu CL, Li JJ, Hu DY, Sun YX. Body mass index and the prevalence of prehypertension and hypertension in a Chinese rural population. *Int Med* 2008; **47**: 893–897.
10 Vasan RS, Beiser A, Seshadri S, Larson MG, Kannel WB, D'Agostino RB, Levy D. Residual lifetime risk for developing hypertension in middle-aged women and men—The Framingham Heart Study. *JAMA* 2002; **287**: 1003–1010.
11 Lin SG, Cheng TO, Liu XQ, Mai JZ, Rao XX, Gao XM, Deng HH, Shi ML. Impact of dysglycemia, body mass index, and waist-to-hip ratio on the prevalence of systemic hypertension in a lean Chinese population. *Am J Cardio* 2006; **97**: 839–842.
12 Ueshima H, Zhang XH, Choudhury SR. Epidemiology of hypertension in China and Japan. *J Hum Hyperten* 2000; **14**: 765–769.
13 Yu Z, Nissnien A, Vartiainen E, Yu ZJ, Nissinen A, Vartiainen E, Hu G, Tian HG, Guo ZY. Socio-economic status and serum lipids: across-sectional study in a Chinese urban population. *J Clin Epidemiol* 2002; **55**: 143–149.
14 Wildman RP, Gu DF, Muntner P, Huang GY, Chen JC, Duan XF, He J. Alcohol intake and hypertension subtypes in Chinese men. *J Hypertension* 2005; **23**: 737–743.
15 World Bank *Sharing Rising Income: Disparities in China*. World Bank: Washington, DC, USA 1997.
16 Sun ZQ, Zheng LQ, Detrano R, Zhang DY, Zhang XG, Xu CL, Li J, Liu SS, Li JJ, Hu DY, Sun Y. The accelerating epidemic of hypertension among rural Chinese women: Results from Liaoning Province. *Am J Hypertens* 2008; **21**: 784–788.
17 Rose GA, Blackburn H, Gillum RF, Prinneas RJ. *Cardiovascular survey methods*. World Health Organization: Geneva, 1982.
18 Weiner JS, Lourie JA. *Practical human biology*. Academic Press: London, 1981.
19 Zhou BF. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults—study on optimal cut-off points of body mass index and waist circumference in Chinese adults. *Biomed Environ Sci* 2002; **15**: 83–96.
20 Adams KF, Schatzkin A, Harris TB, Kipnis V, Mouw T, Barbash RB, Hollenbeck A, Leitzmann MF. Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med* 2006; **355**: 763–778.
21 Community intervention trial for smoking cessation (COMMIT): II. Changes in adult cigarette smoking prevalence. *Am J Public Health* 1995; **85**: 193–200.
22 Jousilahti P, Tuomilehto J, Korhonen HJ, Vartiainen E, Puska P, Nissinen A. Trends in cardiovascular disease risk factor clustering in eastern Finland: results of 15-year follow-up of the North Karelia Project. *Prev Med* 1994; **23**: 6–14.
23 Vartiainen E, Paavola M, McAlister A, Puska P. Fifteen-year follow-up of smoking prevention effects in the North Karelia youth project. *Am J Public Health* 1998; **88**: 81–85.
24 BRFSS Questionnaires. 2003 (Accessed 10 January 2008, at http://www.cdc.gov/brfss/questionnaires/pdf-ques/2003brfss.pdf.).
25 Bell AC, Adair LS, Popkin BM. Understanding the role of mediating risk factors and proxy effects in the association between socio-economic status and untreated hypertension. *Soc Sci Med* 2004; **59**: 275–283.
26 Naimi TS, Brown DW, Brewer RD, Giles WH, Mensah G, Serdula MK, Mokdad AH, Hungerford DW, Lando J, Naimi S, Stroup DF. Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults. *Am J Prev Med* 2005; **28**: 369–373.
27 Mukamal KJ, Ding EL, Djousse L. Alcohol consumption, physical activity, and chronic disease risk factors: a population-based cross-sectional survey. *BMC Public Health* 2006; **6**: 118.
28 Gurney K. *An Introduction to Neural Networks*. Routledge: London, 1997.
29 Duh MS, Wallker A, Ayanian JZ. Epidemiologic interpretation of artificial neural network. *Am J Epidemiol* 1998; **14**: 464–471.
30 Grossi E. How artificial intelligence tools can be used to assess individual patient risk in cardiovascular disease: problems with the current methods. *BMC Cardiovascular Disorders* 2006; **6**: 20.
31 Montie JE, Wei JT. Artificial neural networks for prostate carcinoma risk assessment: An overview. *Cancer* 2000; **88**: 2655–2660.
32 Gary LG, Ruyun J. Receiver operating characteristic curve analysis of clinical risk models. *Ann Thorac Surg* 2001; **72**: 323–326.
33 Hou XH. Urban–Rural disparity of overweight, hypertension, undiagnosed hypertension, and untreated hypertension in China. *Asia-Pacific J Public Health Volume* 2008; **20**: 159–169.
34 Zheng L, Sun Z, Zhang X, Xu C, Li J, Hu D, Sun Y. Predictors of progression from prehypertension to hypertension among rural Chinese adults: results from Liaoning Province. *Eur J Cardiovasc Prev Rehabil* 2010; **17**: 217–222.
35 Zhu J, Zhu XD, Liang SX, Xu ZY, Zhao JD, Huang QF, Wang AY, Chen L, Fu XL, Jiang GL. Prediction of radiation induced liver disease using artificial neural networks. *Jpn J Clin Oncol* 2006; **36**: 783–788.
36 Baldassarre D, Grossi E, Buscema M, Intraligi M, Amato M, Tremoli E, Pustina L, Castelnuovo S, Sanvito S, Gerosa L, Sirtori CR. Recognition of patients with cardiovascular disease by artificial neural networks. *Ann Med* 2004; **36**: 630–640.
37 Santos-García G, Varela G, Novoa N, Jimenez MF. Prediction of postoperative morbidity after lung resection using an artificial neural network ensemble. *Artif Intell Med* 2004; **30**: 61–69.
38 Sato F, Shimada Y, Selaru FM, Shibata D, Maeda M, Watanabe G, Mori Y, Stass SA, Imamura M, Meltzer SJ. Prediction of survival in patients with esophageal carcinoma using artificial neural networks. *Cancer* 2005; **103**: 1596–1605.
39 Matsui Y, Egawa S, Tsukayama C, Terai A, Kuwao S, Baba S, Arai Y. Artificial neural network analysis for predicting pathological stage of clinically localized prostate cancer in the Japanese population. *Jpn J Clin Oncol* 2002; **32**: 530–535.
40 Agyemang C, van Hooijdonk C, Wendel-Vos W, Ujcic-Voortman JK, Lindeman E, Stronks K, Droomers M. Ethnic differences in the effect of environmental stressors on blood pressure and hypertension in the Netherlands. *BMC Public Health* 2007; **7**: 147.