
Wenchao Zhang; Linping Wang, MS; Yafei Chen; Fang Tang, MS; Fuzhong Xue, PhD; Chengqi Zhang, PhD

Abstract

Introduction
Research suggests that targeting high-risk, nonhypertensive patients for preventive intervention may delay the onset of hypertension. We aimed to develop a biomarker-based risk prediction model for assessing hypertension risk in an urban Han Chinese population.

Methods
We analyzed data from 26,496 people with hypertension to extract factors from 11 check-up biomarkers. Then, depending on a 5-year follow-up cohort, a Cox model for predicting hypertension development was built by using extracted factors as predictors. Finally, we created a hypertension synthetic predictor (HSP) by weighting each factor with its risk for hypertension to develop a risk assessment matrix.

Results
After factor analysis, 5 risk factors were extracted from data for both men and women. After a 5-year follow-up, the cohort of participants had an area under receiver operating characteristic curve (area under the curve [AUC]) with an odds ratio (OR) of 0.755 (95% confidence interval [CI], 0.746–0.763) for men and an OR of 0.801 (95% CI, 0.792–0.810) for women. After tenfold cross validation, the AUC was still high, with 0.755 (95% CI, 0.746–0.763) for men and 0.800 (95% CI, 0.791–0.810) for women. An HSP-based 5-year risk matrix provided a convenient tool for risk appraisal.

Conclusion
Hypertension could be explained by 5 factors in a population sample of Chinese urban Han. The HSP may be useful in predicting hypertension.

Introduction
Hypertension is a worldwide public health challenge because of its high frequency and concomitant risks of cardiovascular disease and renal disease (1). Many studies have demonstrated that lifestyle modification can prevent high blood pressure, providing a rationale for the identification of high-risk participants so that early lifestyle intervention strategies can be implemented to prevent hypertension (2–4). In recent years, researchers have established hypertension prediction models for different populations, including Americans (5–8), Iranians (9), and Chinese (10,11). The 2 studies of a Chinese population prediction model (10,11) had areas under the curve (AUCs) that ranged from 71.6% to 73.5%. Inability to incorporate enough laboratory biomarkers and existing recall bias from questionnaire variables limited the effect of the 2 prediction models. This study corrects these limitations. Prevalence, awareness, treatment, and control of hypertension in China, a developing country, differs from that of developed countries (12,13). The purpose of this study was to develop a biomarker-based risk-prediction model for hypertension in a population of urban Han Chinese adults.
Methods

Of 95,785 people aged 18 to 88 years who received annual medical examinations from 2005 through 2010 at the Center for Health Management of Shandong Provincial QianFoShan Hospital and Shandong Provincial Hospital, 26,496 were diagnosed with hypertension at their first check-up. Of 69,289 people without hypertension at baseline, 17,471 (10,239 men and 7,232 women) who received annual clinical and laboratory examinations were selected as a 5-year follow-up cohort.

We selected 11 biomarkers to analyze from the routine health check-ups. They were body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), fasting blood glucose (FBG), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), hemoglobin (Hb), hematocrit (HCT), white blood cell count (WBC), lymphocyte count (LC), and neutrophil granulocyte count (NGC). These biomarkers have been previously associated with hypertension: BMI (14–16), SBP and DBP (17), FBG (18), TG (16), HDL-C (19,20), Hb and HCT (21), and WBC, LC, and NGC (22).

We measured the height and weight of participants who were wearing light clothing and no shoes. BMI was calculated as weight (kg) divided by height squared (m$^2$). SBP and DBP were measured using Omron HEM-907 (QuickMedical) by the cuff-oscillometric method in the right arm of seated participants after a 5-minute rest period. Two measurements were taken, and the 2 BP values were averaged. Peripheral blood samples were obtained in the morning after a 12-hour fast to measure the following biomarkers: FBG, TG, HDL-C, Hb, HCT, WBC, LC, and NGC. The study was approved by the Ethics Committee of School of Public Health, Shandong University, and written informed consent was obtained from all participants.

Hypertension was defined as diastolic blood pressure of 90 mm Hg or more, systolic blood pressure of 140 mm Hg or more, or reported use of medication known to treat hypertension. Age and biomarkers were presented by mean (standard deviation), and Student’s t test was used to distinguish between participants with and without baseline hypertension.

First, to eliminate multicollinearity between the routine check-up biomarkers and to extract risk-related factors of hypertension from them, we used factor analysis with principal component algorithm and varimax rotation from correlation matrix. The criteria for retaining factors were set up as eigenvalue of higher than 1. Further analytical interpretation used biomarkers that share a factor loading of at least 0.50. Second, on the basis of the cohort design, the Cox proportional hazards regression model was built between the hazard function of hypertension and the extracted latent factors:

$$h_i(t) = h_0(t)\exp(\beta_0 \text{age}_i + \beta_1 F_1 + \beta_2 F_2 + \ldots + \beta_k F_k)$$

where $h_i(t)$ is the hazard rate for the $i^{th}$ subject at time $t$, and $h_0(t)$ is the baseline hazard at time $t$. Then, a hypertension synthetic predictor (HSP) was derived by weighting each factor with its regression coefficients using the formula

$$\text{HSP} = \beta_1 F_1 + \beta_2 F_2 + \ldots + \beta_k F_k.$$ After that, we calculated an HSP for each participant. Third, on the basis of the cohort design, the Cox proportional hazards regression model was built again between the hazard function of hypertension and the calculated HSP:

$$h_i(t) = h_0(t)\exp(\theta_{\text{age}} + \theta_{\text{HSP}})$$

The predictive probability of hypertension at year $t$ was calculated by the following formula:

$$P(t) = 1 - S(t) = 1 - \exp(-\int_0^t h_i(u) \exp(\theta) \, du)$$

where $\theta = \theta_{\text{age}} + \theta_{\text{HSP}}$.

Fourth, we used MedCalc software (23) for analysis of receiver operator characteristics (ROC) curve, sensitivity, specificity, and significance ($P$ value) to evaluate the prediction effect. Finally, for each participant from the 5-year cohort study, we calculated relative absolute risk (RAR) using the following equation:

$$\text{RAR} = \frac{P_j(t)}{P_j(t)}$$

where $P_j(t)$ known as absolute risk (AR), was the probability of hypertension at year $t$, in which $j$ denoted the participant’s age. $P_j(t)$ signified the average probability of hypertension at year $t$ in $j^{th}$ age, which can be calculated by the following model:

$$\bar{P}_j(t) = 1 - \bar{S}_j(t) = 1 - \exp(-\int_0^t h_j(u) \exp(\bar{\theta}_j) \, du)$$

where $\bar{\theta}_j = \theta_{\text{age}} + \theta_{\text{HSP}}$, $\bar{HSP}_j$ was the mean of HSP in $j^{th}$ age.
All data analyses in this study were conducted for both men and women. We used ArcGIS 9.1 (Esri) to depict the HSP-based 5-year risk matrix for hypertension risk appraisal. All statistical analyses were performed using SAS version 9.2 (SAS Institute, Inc), and significance was set at $P < .05$.

**Results**

In our study, 26,496 of 95,785 participants had hypertension at baseline, a prevalence of 27.7% (32.6% for men and 19.5% for women). Although hypertension prevalence increased with age in both men and women (Figure 1), it was higher in men than women before age 60 and was similar after age 60. Of the 3,793 participants (2,894 men and 899 women) who did not have hypertension at baseline but had hypertension at the end of year 5, the cumulative incidence was 21.7% (3,793 of 17,471). We calculated the distribution of age and 11 biomarkers among participants with and without baseline hypertension (Table 1), and all variables differed significantly for participants with and without baseline hypertension. Results of the analysis were used to create a correlation matrix for the 11 biomarkers (Appendix A).

After exploring factor analysis (EFA), 5 latent factors were extracted from 11 biomarkers. Combined with explained variance and cumulative variance, factor loadings by principal component analysis with varimax rotation (Table 2). Five latent risk-related factors could explain 72.21% of total variance for men and 72.47% for women: inflammatory factor (IF), blood viscosity factor (BVF), insulin resistance factor (IRF), blood pressure factor (BPF), and lipid resistance factor (LRF). Of the 5 factors, IF was contributed by WBC and LC and NGC, BVF by Hb and HCT, IRF by FBG and TG, BPF by SBP and DBP, and LRF by BMI and HDL-C.

ROC curves for hypertension prediction models are in Figure S1 (Appendix B). The AUC was up to 75.5% for men and 80.1% for women (Figure 2, graphs A1 and A2 for men and graphs B1 and B2 for women), and was 75.5% and 80.0% after tenfold cross validation. These matrices provide a convenient tool for hypertension prediction in clinical and health management. For example, if a man aged 40 came to a hospital for a checkup, and 11 routine health check-up biomarkers (BMI, SBP, DBP, FBG, TG, HDL-C, Hb, HCT, WBC, LC, NGC) were tested, his HSP could be calculated using the formula in Figure 2. After that, we find his absolute risk (AR) and RAR in (graphs A1 and A2) through his age and his HSP. AR shows his predictive probabilities for hypertension are more than 5, and RAR shows his hypertension risk compared with his peers (people aged 40).
Discussion

In our study, the prevalence of hypertension was higher among men (32.6%) than women (19.5%) at baseline. However, hypertension prevalence changes with age. Hypertension prevalence rises more steeply in aging women than in men, perhaps because of hormonal changes during menopause (24–26).

Distribution of age and 11 routine health check-up biomarkers of participants with and without baseline hypertension differed significantly. After factor analysis, 5 latent factors were extracted from 11 biomarkers, not only eliminating the multicollinearity, but also explaining the specific pathogenesis of hypertension. The 5 factors were used to predict hypertension in the following prediction model. The 5 factors are the inflammatory factor (IF), blood viscosity factor (BVF), insulin resistance factor (IRF), blood pressure factor (BPF), and lipid resistance factor (LRF) in both men and women, according to our analysis. The cumulative explained variances of the 5 latent factors were 72.21% for men and 72.47% for women. IF and BVF in particular were identified as the key factors for the variation of hypertension (IF contributes 23.54% for men and 24.64% for women; BVF, 15.78% for men and 15.58% for women). Similar results have been found in other studies. Evidence from human and animal studies suggests that inflammation leads to the development of hypertension and that oxidative stress and endothelial dysfunction are involved in the inflammatory cascade (27). The elevation in blood viscosity could increase peripheral resistance and play a role in the pathogenesis of essential hypertension (28–30).

Two hypertension prediction models have been developed in Chinese populations. Although the power of these prediction models (AUC range: 71.6%–73.5%) was acceptable, their risk algorithm and visualization of risk assessment still had room for improvement. Our HSP-based prediction model had better prediction effect (ROC, 75.5% for men and 80.1% for women; ROC after tenfold cross validation, 75.5% for men and 80.0% for women.). One reason our results were more accurate is because we used laboratory biomarkers rather than questionnaire variables to avoid recall bias. Another reason is that we used both factor analysis and the Cox model, producing a better model for hypertension prediction. Finally, we developed a risk matrix with AR and RAR to represent risk assessment, which was convenient for practical application (Figure 2). For example, men and women who receive a routine health check-up can learn their AR from the risk matrix. Comparing their results with those of other same-age participants may warn patients of their risk and guide their choice of nonpharmacologic measures to prevent hypertension. A limitation to our study was that all participants were urban Han Chinese adults; therefore, our results may not be generalizable to other populations. More validation studies for these prediction models are needed.

This is the first hypertension prediction model developed for urban Han Chinese adults. Because of the large sample size, the estimates from our prediction model were stable, as demonstrated by the tenfold cross validation. Physicians can use the HSP-based 5-year hypertension risk matrix to measure patients’ risk for hypertension, inform patients of their risks, help them choose appropriate nonpharmacologic measures to prevent hypertension, and aid in clinical counselling and decision making.

Acknowledgments

This work was supported by grants from the National Nature Science Fund (no. 81273082) of China and Science and Technology Development Projects of Shandong Province (2009GG20002035). The funders were not involved in study design, analysis, or interpretation. The contributions of Wenchao Zhang and Linping Wang to this article were equal.

Author Information

Corresponding Author: Chengqi Zhang, PhD, Professor, Health Management Center, Shandong Provincial QianFoShan Hospital, Jinan 250014, China. Telephone: 011-86-531-89268299. Email: chengqizhangsd@126.com.

Author Affiliations: Wenchao Zhang, Yafei Chen, Fuzhong Xue, Department of Biostatistics, School of Public Health, Shandong University, Jinan, China; Linping Wang, Fang Tang, Health Management Center, Shandong Provincial QianFoShan Hospital, Jinan, China.

References


Table 1. Distribution of Age and the 11 Biomarkers for Participants With and Without Baseline Hypertension, China, 2005–2010

<table>
<thead>
<tr>
<th>Variable</th>
<th>Without Baseline Hypertension (n = 69,236)</th>
<th>With Baseline Hypertension (n = 26,478)</th>
<th>P Values&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>41.37 (12.37)</td>
<td>53.62 (14.24)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Body mass index, kg/m&lt;sup&gt;2&lt;/sup&gt;</td>
<td>24.27 (3.59)</td>
<td>26.86 (3.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Systolic blood pressure, mmHg</td>
<td>117.4 (12.53)</td>
<td>149.2 (16.29)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Diastolic blood pressure, mmHg</td>
<td>70.84 (9.24)</td>
<td>86.30 (12.73)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fasting blood-glucose, mmol/L</td>
<td>5.20 (1.11)</td>
<td>5.80 (1.60)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Triglycerides, mmol/L</td>
<td>1.49 (1.32)</td>
<td>2.06 (1.79)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High-density lipoprotein cholesterol, mmol/L</td>
<td>1.33 (0.33)</td>
<td>1.30 (0.36)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hemoglobin, g/L</td>
<td>146.7 (15.51)</td>
<td>151.1 (14.40)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hematocrit, %</td>
<td>43.61 (4.05)</td>
<td>44.74 (3.79)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>White blood cell count, 10&lt;sup&gt;9&lt;/sup&gt;/L</td>
<td>6.46 (1.59)</td>
<td>6.85 (1.68)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lymphocyte, 10&lt;sup&gt;9&lt;/sup&gt;/L</td>
<td>2.07 (0.61)</td>
<td>2.15 (0.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Neutrophil granulocyte, 10&lt;sup&gt;9&lt;/sup&gt;/L</td>
<td>3.84 (1.23)</td>
<td>4.11 (1.26)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>P values were calculated using Student’s t test.
Table 2. Factor Loadings by Principal Component Analysis With Varimax Rotation on 11 Routine Health Check-up Biomarkers in Hypertension Patients, China, 2005–2010

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Inflammatory</th>
<th>Blood Viscidity</th>
<th>IR</th>
<th>BP</th>
<th>Fat Resistance</th>
<th>Inflammatory</th>
<th>Blood Viscidity</th>
<th>BP</th>
<th>IR</th>
<th>Fat Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>0.0630</td>
<td>0.0978</td>
<td>0.3681</td>
<td>0.1534</td>
<td>-0.6950</td>
<td>0.1000</td>
<td>0.0486</td>
<td>0.3131</td>
<td>0.2566</td>
<td>-0.6194</td>
</tr>
<tr>
<td>SBP</td>
<td>0.0527</td>
<td>-0.1677</td>
<td>-0.0245</td>
<td>0.8540       &amp;superscript;</td>
<td>0.0113</td>
<td>0.0419</td>
<td>-0.0574</td>
<td>0.7988 &amp;superscript;</td>
<td>0.1902</td>
<td>0.0234</td>
</tr>
<tr>
<td>DBP</td>
<td>-0.0003</td>
<td>0.3203</td>
<td>0.1370</td>
<td>0.7064 &amp;superscript;</td>
<td>-0.0595</td>
<td>0.0114</td>
<td>0.1837</td>
<td>0.7549 &amp;superscript;</td>
<td>-0.1657</td>
<td>-0.1249</td>
</tr>
<tr>
<td>FBG</td>
<td>0.0342</td>
<td>-0.0994</td>
<td>0.6782 &amp;superscript;</td>
<td>0.0696</td>
<td>-0.0777</td>
<td>0.0785</td>
<td>0.0423</td>
<td>0.0011 &amp;superscript;</td>
<td>0.7670 &amp;superscript;</td>
<td>-0.0739</td>
</tr>
<tr>
<td>TG</td>
<td>0.0836</td>
<td>0.1380</td>
<td>0.7850 &amp;superscript;</td>
<td>-0.0029</td>
<td>0.1003</td>
<td>0.0807</td>
<td>0.0798</td>
<td>0.0356 &amp;superscript;</td>
<td>0.7250 &amp;superscript;</td>
<td>-0.0163</td>
</tr>
<tr>
<td>HDL-C</td>
<td>-0.0650</td>
<td>-0.0171</td>
<td>0.2468</td>
<td>0.0761</td>
<td>0.8323 &amp;superscript;</td>
<td>-0.0431</td>
<td>-0.0120</td>
<td>0.0782</td>
<td>0.0574</td>
<td>0.8778 &amp;superscript;</td>
</tr>
<tr>
<td>Hb</td>
<td>0.0627</td>
<td>0.9574 &amp;superscript;</td>
<td>0.0450</td>
<td>0.0333</td>
<td>-0.0550</td>
<td>0.0639</td>
<td>0.9705 &amp;superscript;</td>
<td>0.0542</td>
<td>0.0920</td>
<td>-0.0403</td>
</tr>
<tr>
<td>HCT</td>
<td>0.1242</td>
<td>0.9472 &amp;superscript;</td>
<td>-0.0234</td>
<td>0.0332</td>
<td>-0.0425</td>
<td>0.1146</td>
<td>0.9651 &amp;superscript;</td>
<td>0.0674</td>
<td>0.0640</td>
<td>-0.0138</td>
</tr>
<tr>
<td>WBC</td>
<td>0.9906 &amp;superscript;</td>
<td>0.0718</td>
<td>0.0431</td>
<td>0.0355</td>
<td>-0.0512</td>
<td>0.9872 &amp;superscript;</td>
<td>0.0747</td>
<td>0.0170</td>
<td>0.0626</td>
<td>-0.0957</td>
</tr>
<tr>
<td>LC</td>
<td>0.6468 &amp;superscript;</td>
<td>0.0577</td>
<td>0.0850</td>
<td>-0.0199</td>
<td>-0.0306</td>
<td>0.6524 &amp;superscript;</td>
<td>0.0637</td>
<td>0.0837</td>
<td>0.1971</td>
<td>0.0990</td>
</tr>
<tr>
<td>NGC</td>
<td>0.8548 &amp;superscript;</td>
<td>0.0544</td>
<td>0.0036</td>
<td>0.0589</td>
<td>-0.0454</td>
<td>0.8577 &amp;superscript;</td>
<td>0.0571</td>
<td>-0.0277</td>
<td>-0.0327</td>
<td>-0.1728</td>
</tr>
<tr>
<td>% Variance explained</td>
<td>0.2354</td>
<td>0.1578</td>
<td>0.1260</td>
<td>0.1060</td>
<td>0.0969</td>
<td>0.2464</td>
<td>0.1558</td>
<td>0.1237</td>
<td>0.1037</td>
<td>0.0951</td>
</tr>
<tr>
<td>Cumulative variance</td>
<td>0.2354</td>
<td>0.3932</td>
<td>0.5192</td>
<td>0.6252</td>
<td>0.7221</td>
<td>0.2464</td>
<td>0.4022</td>
<td>0.5259</td>
<td>0.6296</td>
<td>0.7247</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; BP, blood pressure; DBP, diastolic blood pressure; FBG, fasting blood glucose; Hb, hemoglobin; HCT, hematocrit; HDL-C, high-density lipoprotein cholesterol; IR, insulin resistance; LC, lymphocyte; NGC, neutrophil granulocyte; SBP, systolic blood pressure; TG, triglycerides; WBC, white blood cell count.

* Factor loading >0.5.
Appendix A

Correlation Matrix Between 11 Biomarkers, Urban Han Population, China, 2005–2010. This file is available for download as a Microsoft Word document at http://www.cdc.gov/pcd/issues/2015/docs/15_0192_AppendixA.docx [DOC – 41 KB].
Appendix B