

## **Predicting the Risk of Coronary Heart Disease and Diabetes in Taiwan**

Hsing-Yi Chang, Ching-Yu Huang, Hsin-Ling Fang, Wen-Han Pan

Institute of Population Health Sciences, National Health Research Institutes,  
Miaoli, Taiwan

An accurate risk prediction model can be used to prevent diseases and promote health. Recently, many methods in selecting models have been developed. Discrimination and calibration are two major elements for evaluating the performance of prediction models. Measures of the discrimination performance of models are derived from Harrell's C statistic, and the calibration abilities are derived from Hosmer-Lemeshow test. We used the Nutrition and Health Survey in Taiwan (NAHSIT, 1993-1996) linked to National Health Insurance database and mortality records (1996-2006) to construct the disease risk prediction models for coronary heart disease, diabetes, and chronic kidney disease. Results show that age, systolic blood pressure, triglyceride, uric acid, and BMI are major risk factor for CHD, and age, fasting glucose, triglyceride, and the waist circumferences are major risk factors for diabetes. Our models can predict the probability of having CHD or diabetes in 10 years given risk factors. The values of C statistic and H-L test for CHD risk prediction model are 0.67 and 13.68 for men, and 0.68 and 10.21 for women. The values of C statistic and H-L test for diabetes risk prediction model are 0.7 and 10.65 for men, and 0.73 and 22.15 for women. That indicates the models fit the data well. We will use another data set to validate the results.

Key words: risk prediction model; Cox proportional model; discrimination; calibration; Harrell's C statistic; Hosmer-Lemeshow test

## **Background**

As the availability of biomarkers grows rapidly, disease risk prediction models are also increasing. Diabetes has been among the top 5 leading causes of death in Taiwan. The mortality has been much higher than most other countries. The methods in selecting models have also been developed. The problem of coronary heart disease is worse than diabetes. Therefore, developing risk prediction models for them is crucial for Taiwan.

An accurate risk prediction model can be used to prevent diseases and promote health. Recently, many methods in selecting models have been developed. Discrimination and calibration are two major elements for evaluating the performance of prediction models. Measures of the discrimination performance of models are derived from Harrell's C statistic, and the calibration abilities are derived from Hosmer-Lemeshow test. The purpose of this study is to develop risk prediction models for coronary heart disease and diabetes in Taiwan.

## **Material and Method**

We used the Nutrition and Health Survey in Taiwan (NAHSIT, 1993-1996) linked to National Health Insurance database and mortality records (1996-2006) to construct the disease risk prediction models. Coronary heart disease was identified with ICD-9: 410-414, where as diabetes was identified as ICD-9: 250. The date of diagnosed was used as the event time. Subjects free of disease were used and followed to the end of 2007. We constructed coronary heart disease (CHD) and diabetes prediction models by using the Cox proportional model [1]. Measures of the discrimination performance of models are derived from Harrell's C statistic [2], and the calibration abilities are derived from Hosmer-Lemeshow test.

## **Results**

After exclusion of those having the disease, with missing values of data or outlier, 1481 and 1451 participants were included in the study for CHD and diabetes respectively. Table 1 shows the characteristics of the study subjects. The average age of those developed the diseases was around 56 years, significantly older than those without the disease during the follow-up period. They also had higher blood pressures, total

cholesterol, triglycerides, low density lipoprotein (LDL), fasting glucose, and waist circumferences than their counterparts (Table 1).

Table 2 shows the selected risk factors for each disease by gender. Age, body mass index, systolic blood pressure, triglycerides, and uric acid are predictors for CHD. Age, waist circumferences, triglycerides and fasting glucose are predictors for diabetes. The test statistics are listed in Table 3. The C statistics are around 0.7 and the chi-squares are less than 15 for the models. We also plotted the predicted and the observed proportion in Figure 1.

## Discussion

The selected predictors are routinely collected in health check-up. Once the health check-up is performed, physicians can provide the probability of getting these diseases to the subjects. It should serve as a prevention scale to alert the subject in changing lifestyles. The selected factors are related to metabolic disorder or obesity. Lifestyle modification would alter the chance of getting the disease.

In order to check the robustness of the model performance, we will use another data set to validate these results. The final models may help to target Taiwan population at risk of developing diseases and have the potential in suggesting people to modify their lifestyles thus to prevent these diseases.

## References

1. D'Agostino RB, Griffith JL, Schmidt CH, Terrin N. *Measures for evaluating model performance*. Proceedings of the Biometrics Section. American Statistical Association, Biometrics Section: Alexandria, VA, 1997; 253-258.
2. Harrell FE et al. *Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors*. *Statistic in Medicine* 1996; 15: 361-387.

Table 1. Characteristics of the study subjects.

Basic Characteristics in Study Samples with and without Diseases (CHD and Diabetes)										
	CHD (n=468)		NO CHD (m=1013)		P*	Diabetes (n=392)		NO Diabetes (n=1059)		P*
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Age, y	56.9	8.9	46.4	9.6	<.0001	55.9	9.3	49.8	9.7	<.0001
Body mass index, kg/m <sup>2</sup>	24.8	3.4	24.0	3.4	<.0001	25.4	3.8	23.8	3.4	<.0001
Systolic blood pressure, mm Hg	135.2	21.5	124.5	18.0	<.0001	133.7	20.7	124.3	18.6	<.0001
Diastolic blood pressure, mm Hg	83.7	14.0	80.8	11.9	0.0001	83.9	13.6	80.7	12.2	<.0001
Total cholesterol, mg/dL	204.0	42.8	196.2	35.1	0.0007	203.8	40.9	196.3	37.4	0.0015
Triglyceride, mg/dL	135.9	83.7	116.1	74.1	<.0001	140.6	84.1	111.2	69.4	<.0001
HDL cholesterol, mg/dL	53.9	17.1	56.8	18.2	0.0036	53.7	16.5	57.3	18.3	0.0007
LDL cholesterol, mg/dL	123.0	35.9	116.3	35.6	0.003	122.0	37.9	116.7	36.1	0.0177
Uric acid, mg/dL	6.4	1.9	6.0	1.6	<.0001	6.7	2.0	5.9	1.6	<.0001
Fasting glucose, mg/dL	107.8	40.3	96.4	21.5	<.0001	94.8	11.2	92.1	8.8	<.0001
Waist circumferences	82.6	9.3	78.7	9.2	<.0001	83.3	9.6	78.0	9.0	<.0001

SD= standard deviation, HDL= high-density lipoprotein; LDL= low-density lipoprotein.

Table 2. Selected risk factors in predicting diabetes in men and women.

Estimated Coefficient, Standard Error from the Cox Regression for the Prediction Models of the Three Diseases					
		Men		Women	
Risk factors		Estimated Coefficient	SE	Estimated Coefficient	SE
<b>CHD model (Men/Women)</b>	Age, y	0.0593	0.0072	0.0603	0.0078
	Body mass index, kg/m <sup>2</sup>	0.0031	0.0212	0.0371	0.0194
	Systolic blood pressure, mm Hg	0.0112	0.0034	0.0077	0.0033
	Triglyceride, mg/dL	0.0016	0.0008	0.0012	0.0008
	Uric acid, mg/dL	0.0385	0.0371	0.0719	0.0193
<b>Diabetes model (Men/Women)</b>	Age, y	0.0541	0.0076	0.0381	0.0078
	Waist circumferences	0.0241	0.0076	0.0473	0.0082
	Fasting glucose, mg/dL	0.0076	0.0071	0.0241	0.0077
	Triglyceride, mg/dL	0.0027	0.0008	0.0017	0.0008
<b>CKD model (Men/Women)</b>	Age, y	0.0652	0.0083	0.0786	0.0102
	Diastolic blood pressure, mm Hg	0.0144	0.0057	0.0035	0.0074
	Triglyceride, mg/dL	0.0015	0.0009	0.0016	0.0010
	Uric acid, mg/dL	0.1292	0.0416	0.0807	0.0518

Table 3. Test statistics for CHD and diabetes models

Summary statistics for the performance of models				
Men		Women		
Model	Discrimination Harrell's C statistic	Hosmer-Lemeshow test	Discrimination Harrell's C statistic	Hosmer-Lemeshow test
<b>CHD</b>	0.69	16.64	0.71	11.03
<b>Diabetes</b>	0.68	14.18	0.73	14.82

Figure 1. The predicted and observed events.

