

Supplementary material

Shangguan C, Yu L, Liu G, et al. Risk assessment of chronic obstructive pulmonary disease using a Bayesian network based on a provincial survey. Pol Arch Intern Med. 2021; 131: 345-355. doi:10.20452/pamw.15867

Please note that the journal is not responsible for the scientific accuracy or functionality of any supplementary material submitted by the authors. Any queries (except missing content) should be directed to the corresponding author of the article.

Table S1. Blacklists of the BN

From	To	From	To	From	To	From	To
Smoke		Age		Age		Coalbio	Income
Wheezing		Smoke		Smoke		Venti	Region
Edu		Wheezing		Wheezing			
Air_sat		Edu		Air_sat			
CHD		Air_sat		CHD			
Income		CHD		Income			
BMI		Income		BMI			
COPD	Age	BMI	Sex	COPD	Edu		
sm14		COPD		sm14			
Sex		sm14		Sex			
Venti		Venti		Venti			
Dust		Dust		Dust			
Gas		Gas		Gas			
Region		Region		Coalbio			
Coalbio		Coalbio		Cough			

Cough		Cough					
--------------	--	--------------	--	--	--	--	--

Note :

CHD: Coronary heart disease; Coalbio: exposure to coal or biomass fuel during cooking; sm14: environmental tobacco smoke exposure between 14 years old; Air_sat: the satisfaction of the ambient air quality; Venti: Ventilation of the kitchen; Edu: education level.

The age or sex could not be affected by the other features. For education, the listed variables are mostly current features that are not able to affect the education levels determined decades before.

Table S2. Whitelist of the BN

from	to
Age	COPD
Smoke	COPD
Sex	COPD
Gas	COPD
Dust	COPD
COPD	Cough
COPD	Wheezing
Income	COPD
BMI	COPD

Note :

Age: Age group; Gas: Occupational exposure to hazardous gas; Dust: Occupational exposure to dust; Wheezing: Wheezing like sound during sickness.

Table S3. Comparison of AUC and accuracy between BN and accuracy during cross-validation

Folds	AUC		Accuracy	
	BN	Logistic regression	BN	Logistic regression
1	0.849	0.776	0.876	0.834
2	0.86	0.778	0.876	0.837
3	0.892	0.773	0.893	0.864
4	0.823	0.776	0.846	0.825
5	0.847	0.764	0.859	0.840

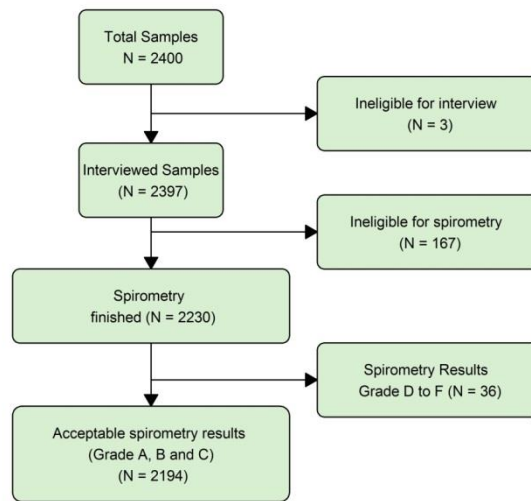


Figure S1. The study process. The participants were randomly sampled in 4 cities of Liaoning Province, China. The participants were interviewed, and they also underwent pre-bronchodilator and post-bronchodilator spirometry, if qualified. The results were examined by the staff, and the qualified observations were included in the multivariate analysis. The Bayesian Network was constructed based on a priori that consisted of multivariate logistic results and previous evidence. Validations were followed to validate the model, and an area under the curve (AUC) of at least 0.75 was considered acceptable.

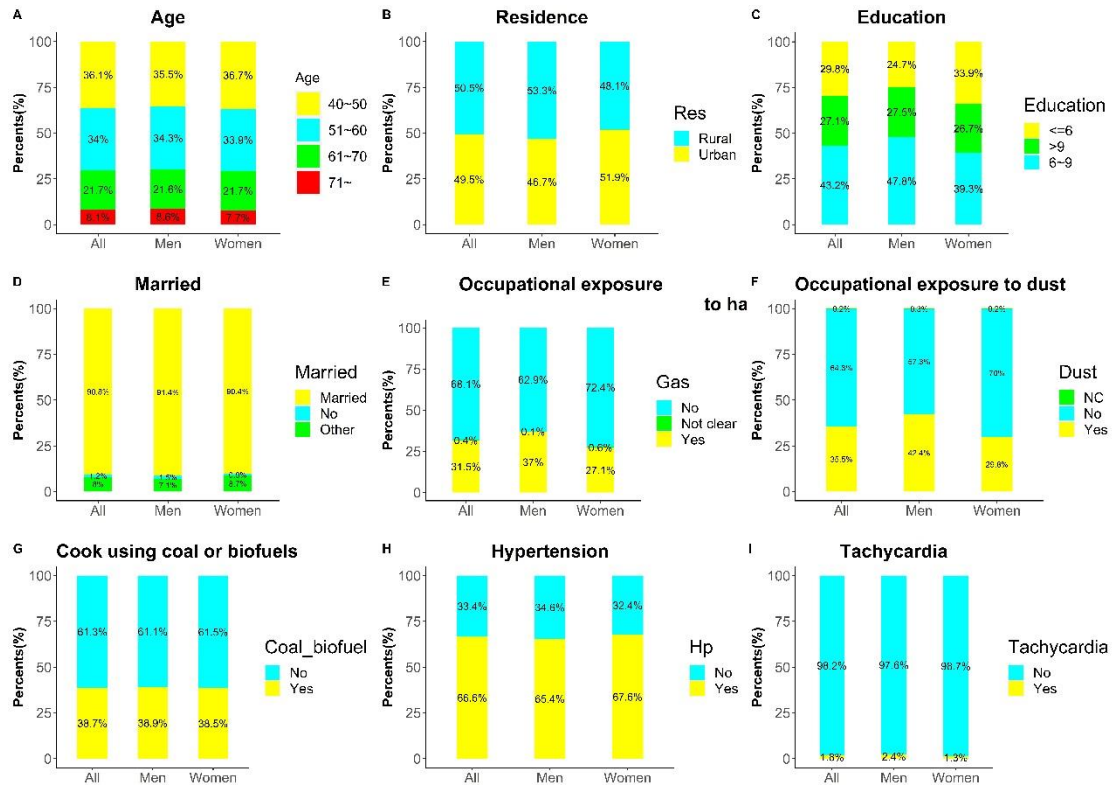


Figure S2. Barchart of composition of some principle variables (A). (A): Age group; (B): Resident areas, including rural areas and urban areas; (C): Education level; (D): Marrital status; (E): Occupational Exposure to Hazardous Gas; (F) Occupational Exposure to Dust; (G): Using Biomass fuel or coal for cooking; (H): Hyperteion or not; (I): Tachycardia or not.

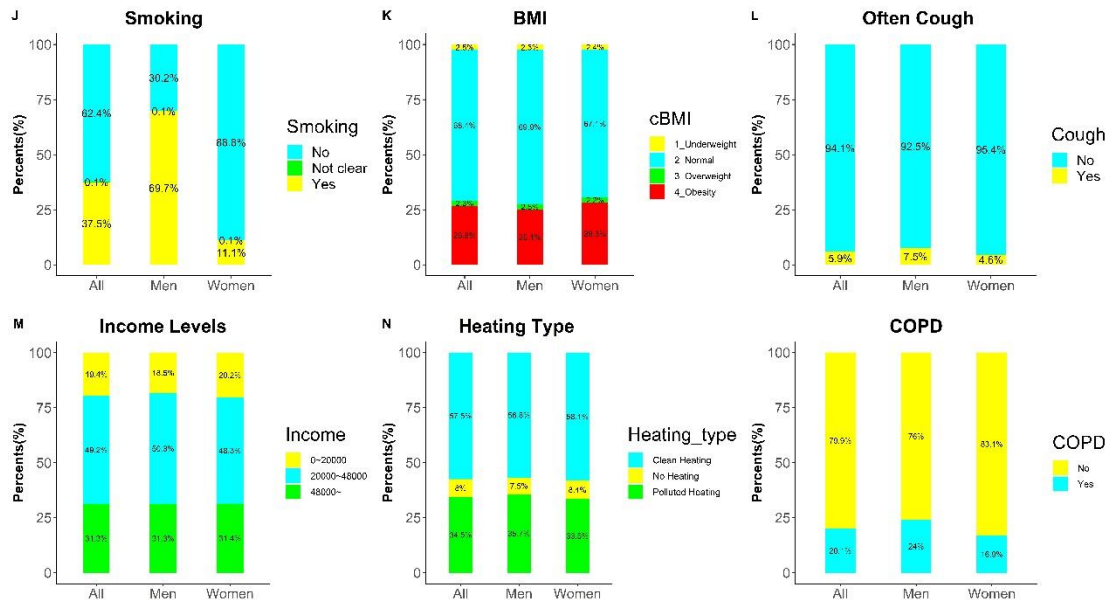


Figure S3. Barchart of composition of some principle variables (B). (J): Ever smoking; (K): BMI group, including underweight, normal, overweight and obesity; (L): Frequent cough or not; (M): Income level group, including low (less than 20,000 RMB), middle (20,000-48,000) and high (more than 48,000) group; (N): Heating type, including clean energy (electry or gas) and polluted energy (coal or biomass fuel); COPD: Chronic obstructive pulmonary disease or not.

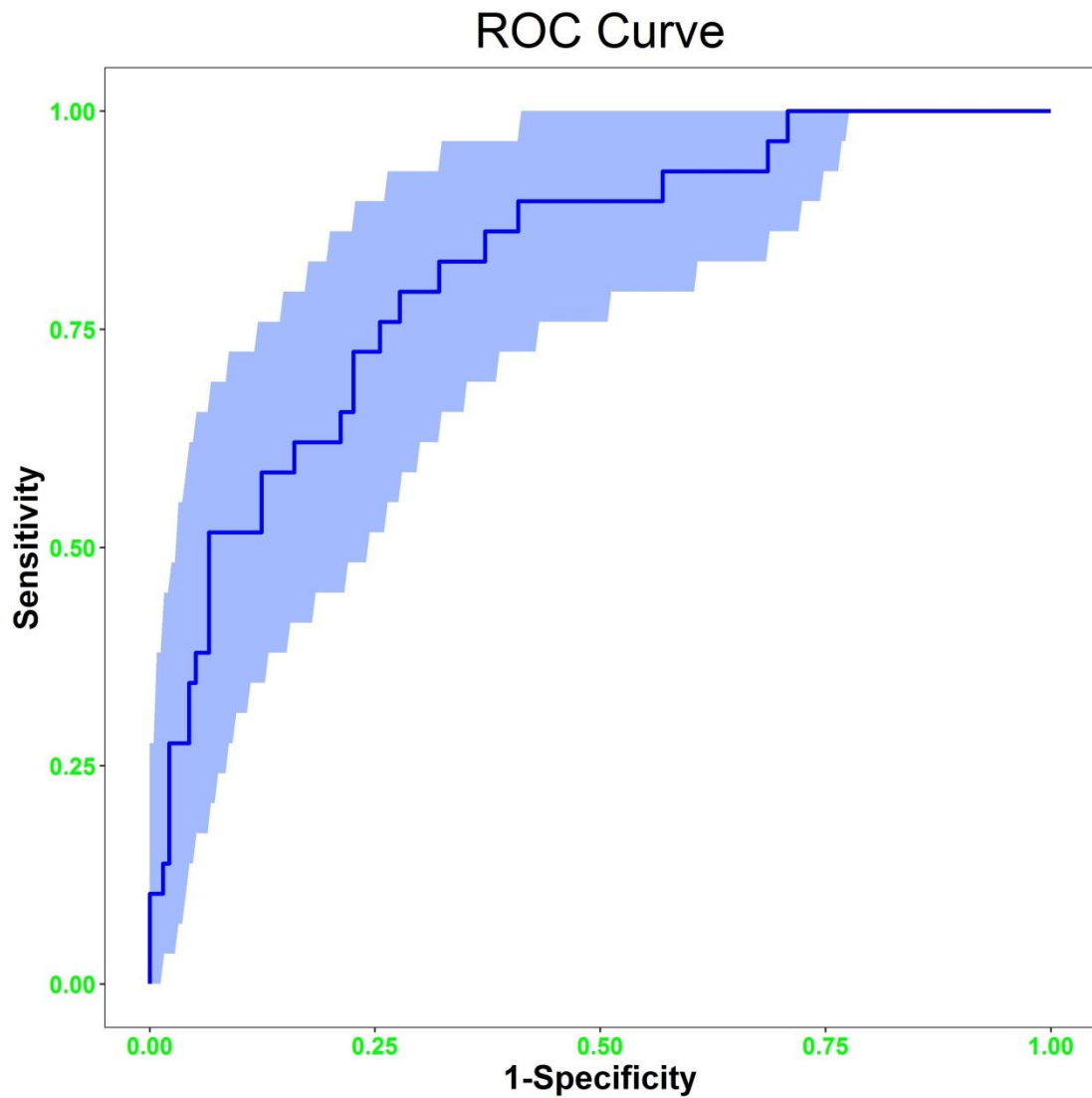


Figure S4. The receiver operating characteristic curves created during the external validations. The Bayesian Network model had a relatively good performance in external validation.