

**Table S1.** Prediction Probability Calculation

HCHS Model:

$$1/[1 + \exp[-[-10.2561 + 0.0655 \times \text{Age} + 0.1391 \times \text{BMI} + 0.7006 \times (1 \text{ if male, } 0 \text{ if female}) + 0.9481 \times (1 \text{ if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) + 0.1012 \times (1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise})]]]$$

Model 1:

$$1/[1 + \exp[-[-8.438927 + 0.054307 \times \text{Age} + 0.123250 \times \text{BMI} + 0.856157 \times (1 \text{ if male, } 0 \text{ if female}) + 0.710687 \times (1 \text{ if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.427722 \times (1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise})]]]$$

Model 2:

$$1/[1 + \exp[-[-11.715931 + 0.056726 \times \text{Age} + 0.095412 \times \text{BMI} + 0.343521 \times (1 \text{ if male, } 0 \text{ if female}) + 0.673186 \times (1 \text{ if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.403381 \times (1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise}) + 0.105496 \times \text{Neck size} + 0.603378 \times (1 \text{ if restless or very restless sleep, } 0 \text{ if otherwise})]]]$$

Model 3:

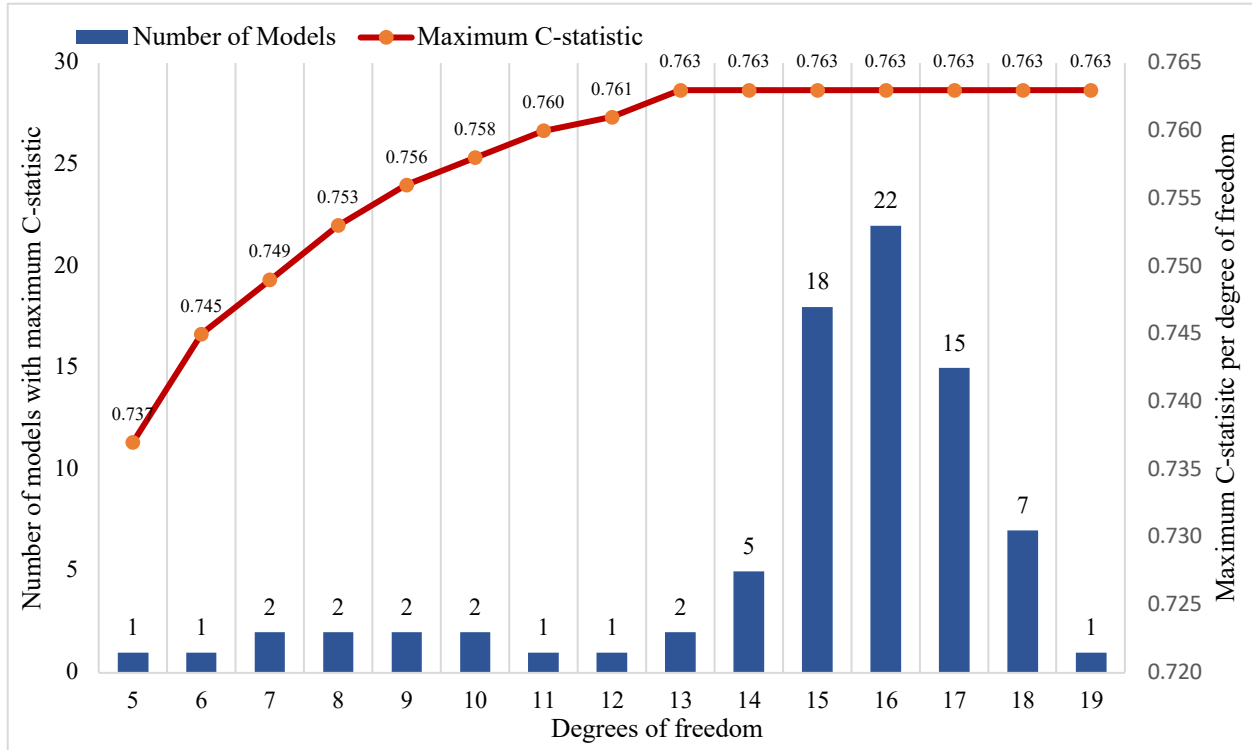
$$1/[1 + \exp[-[-52.862590 + 2.090942 \times \text{Age} - 0.032839 \times \text{Age}^2 + 0.000173 \times \text{Age}^3 + 0.099251 \times \text{BMI} + 0.379448 \times (1 \text{ if male, } 0 \text{ if female}) + 0.628787 \times (1 \text{ if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.456127 \times (1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise}) + 0.104075 \times \text{Neck size} + 0.604672 \times (1 \text{ if restless or very restless sleep, } 0 \text{ if otherwise})]]]$$

Model 4:

$$1/[1 + \exp[-[-43.195210 + 2.073843 \times \text{Age} - 0.032705 \times \text{Age}^2 + 0.0001732 \times \text{Age}^3 + 0.097365 \times \text{BMI} + 0.250963 \times (1 \text{ if male, } 0 \text{ if female}) + 0.622340 \times (1 \text{ if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.496796 \times (1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise}) - 0.363483 \times \text{Neck size} + 0.006025 \times \text{Neck size}^2 + 0.832286 \times (1 \text{ if restless or very restless sleep, } 0 \text{ if otherwise}) + 0.584948 \times (1 \text{ if observed stopped breathing} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.386972 \times (1 \text{ if trouble falling asleep} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.377969 \times (1 \text{ if CESD} - 20 \geq 16, 0 \text{ if otherwise})]]]$$

Model 5:

$$1/[1 + \exp[-[-24.585590 + 1.677163 \times \text{Age} - 0.026769 \times \text{Age}^2 + 0.000144 \times \text{Age}^3 + 0.073410 \times \text{BMI} - 0.001598 \times \text{BMI}^2 + 0.4532339 \times (1 \text{ if male, } 0 \text{ if female}) + 0.626873 \times (1 \text{ if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.483827 \times (1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise}) - 0.472079 \times \text{Neck size} + 0.007199 \times \text{Neck size}^2 + 0.275387 \times \text{Sleepiness} + 0.735190 \times (1 \text{ if restless or very restless sleep, } 0 \text{ if otherwise}) + 0.661046 \times (1 \text{ if observed stopped breathing} \geq 3 \text{ times a week, } 0 \text{ if otherwise}) - 0.797111 \times (1 \text{ if Sleepiness and Male, } 0 \text{ if otherwise}) - 0.429219 \times (1 \text{ if CESD} - 20 \geq 16, 0 \text{ if otherwise}) + 0.100862 \times \text{sleep duration} - 0.282648 \times \text{Waist size} + 0.003090 \times \text{Waist size}^2 - 0.000010 \times \text{Waist size}^3]]]$$



**Figure S1.** Number of models with maximum C-statistic per degrees of freedom

**Notes:** X-axis represents the number of degrees of freedom in the model which ranges from 5 to 19. Blue bars were used to mark the number of models with the same degrees of freedom and maximum C-statistic, whose values correspond to the Y-axis on the left side. The red line above the bars was connected by the orange dots showing the maximum C-statistic per degrees of freedom (Y-axis on the right side). For example, there are 22 models with 16 degrees of freedom that have a C-statistic of 0.763.

**Table S2.** C-statistic of 5 models retained after Bootstrap in the final step of 5-fold cross-validation. C-statistics are provided for the full analytic data set and averaged in the 5-fold cross-validation.

Prediction model	Number of measures	Degrees of freedom	Complete dataset (C-statistic and 95% CI)	5-fold cross-validation (average C-statistic)
				Testing datasets
<b>Model 1:</b> Age, BMI, Male, Snoring	4	5	0.737 (0.696, 0.778)	0.723
<b>Model 2:</b> Model 1 + Restless or Very Restless Sleep + Neck Size	6	7	0.749 (0.708, 0.789)	0.739
<b>Model 3:</b> Model 2 + Witnessed Apneas + High Depressive Symptoms + Average Sleep Duration	9	10	0.758 (0.718, 0.798)	0.746
<b>Model 4:</b> Model 3 + History of heart diseases + Trouble falling asleep + Multiple awakenings at night + Naps + Waist Size + Waist Size <sup>2</sup>	15	16	0.763 (0.724, 0.803)	0.740

**Table S3.** Predictive properties by model and cut-off.

Method	C-Statistic (95% CI)	Cutoffs	High Risk (%)	Accuracy (%)	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
Model 1	0.737 (0.696, 0.778)	Optimal (probability $\geq$ 0.26)	37%	70%	0.65 (0.55, 0.72)	0.71 (0.59, 0.76)	0.36
Model 2	0.749 (0.708, 0.789)	Optimal (probability $\geq$ 0.24)	41%	69%	0.70 (0.59, 0.77)	0.69 (0.56, 0.73)	0.39
Model 3	0.758 (0.718, 0.798)	Optimal (probability $\geq$ 0.29)	31%	74%	0.60 (0.50, 0.67)	0.78 (0.66, 0.82)	0.38
Model 4	0.763 (0.724, 0.803)	Optimal (probability $\geq$ 0.28)	35%	72%	0.64 (0.53, 0.71)	0.75 (0.62, 0.79)	0.39

Note: The true prevalence of moderate or sleep apnea (AHI 4%) in the complete dataset was 24%.