

Supplementary Table 1 General information about the CNDMDS and NHANES datasets

Characteristics	Derivation dataset-	Validation dataset-NHANES ^b				
	CNDMDS ^a	Mexican American	Other Hispanic	None-Hispanic White	None-Hispanic Black	Non-Hispanic Asian
Study period	2007-2008	2011-2012	2011-2012	2011-2012	2011-2012	2011-2012
Location	14 provinces across China	U.S.	U.S.	U.S.	U.S.	U.S.
Study population	40381 participants	166 Mexican American	160 other Hispanic	631 non-Hispanic White	337 non-Hispanic Black	231 Asian
Proportion of participants with prediabetes	16.9% (6810/40381)	22.9% (38/166)	23.8% (38/160)	22.3% (141/631)	22.3% (75/337)	17.3% (40/231)

^aStudy participants in the CNDMDS for model derivation included 15848 men and 24533 women who were aged ≥ 20 years old.

^bValidation dataset included Hispanic and non-Hispanic populations in NHANES, who were aged between 20 and 80 years old without missing information on key indicators.

Supplementary Table 2 Unadjusted association between candidate predictors and prediabetes

Variables	Odds ratio	95% confidence interval (CI)	P value
Age, years	1.040	(1.038-1.043)	< 0.0001
Male	1.014	(0.962-1.070)	0.5983
Middle school or higher education	0.572	(0.540-0.606)	< 0.0001
DM family history	1.335	(1.237-1.439)	< 0.0001
Ever smoker	1.039	(0.978-1.103)	0.2154
Waist circumference \geq normal upper limit ¹	2.355	(2.232-2.485)	< 0.0001
BMI, kg/m ²	1.133	(1.125-1.141)	< 0.0001
SBP, mmHg	1.026	(1.024-1.027)	< 0.0001
DBP, mmHg	1.032	(1.030-1.035)	< 0.0001

¹Waist circumference \geq 90 cm in men and \geq 85 cm in women were considered to exceed the normal waist circumference.

Abbreviations: DM=diabetes mellitus, BMI=body mass index, SBP=systolic blood pressure, DBP=diastolic blood pressure.

Supplementary Table 3 Characteristics of the study populations in the CNDMDS and NHANES

Characteristics	CNDMDS	NHANES				
	N=40381	Mexican American	Other Hispanic	None-Hispanic	None-Hispanic	Non-Hispanic
		N=166	N=160	White N=631	Black N=337	Asian N=231
Age, years	43.8±13.4	40.6±15.2	46.1±16.2	47.6±17.7	45.2±16.4	41.4±15.0
Men	15848 (39.2)	97 (58.4)	85 (53.1)	330 (52.3)	152 (45.1)	125 (54.1)
Middle school or higher education	31175 (77.2)	93 (56.0)	110 (68.8)	529 (83.8)	280 (83.1)	205 (88.7)
DM family history	4736 (11.7)	65 (39.2)	48 (30.0)	189 (30.0)	159 (47.2)	56 (24.2)
Waist circumference ≥ normal upper limit*	11906 (29.5)	78 (47.0)	80 (50.0)	347 (55.0)	185 (54.9)	76 (32.9)
BMI, kg/m ²	23.9±3.6	28.9±5.5	28.4±5.3	28.4±6.1	29.8±7.0	23.8±4.5

Characteristics	CNDMDS	NHANES				
	N=40381	Mexican American	Other Hispanic	None-Hispanic	None-Hispanic	Non-Hispanic
		N=166	N=160	White N=631	Black N=337	Asian N=231
SBP, mm Hg	121±19	116±15	119±16	120±15	124±18	117±16
DBP, mm Hg	78±11	68±11	70±12	70±12	71±13	71±9
Cholesterol, mmol/L	4.7±1.0	4.9±1.0	5.0±0.9	5.0±1.0	4.9±1.0	5.0±0.9
Triglycerides, mmol/L	1.5±1.1	1.4±0.7	1.4±0.7	1.5±1.2	1.0±0.6	1.4±0.9
HDLC, mmol/L	1.3±0.3	1.3±0.3	1.3±0.3	1.4±0.4	1.5±0.4	1.4±0.4
LDLC, mmol/L	2.7±0.8	3.0±0.8	3.1±0.8	3.0±0.9	2.9±0.9	2.9±0.8

Characteristics	CNDMDS	NHANES				
	N=40381	Mexican American	Other Hispanic	None-Hispanic	None-Hispanic	Non-Hispanic
		N=166	N=160	White N=631	Black N=337	Asian N=231
Fasting plasma glucose, mmol/L	5.1±0.6	5.5±0.4	5.5±0.5	5.4±0.5	5.4±0.6	5.4±0.5
Two-hour plasma glucose, mmol/L	6.1±1.6	6.0±1.8	6.1±1.7	5.9±1.7	6.0±1.7	6.1±1.6
Prediabetes, <i>n</i> (%)	6810 (16.9)	38 (22.9)	38 (23.8)	141 (22.3)	75 (22.3)	40 (17.3)

* Waist circumference ≥ 90 cm in men and ≥ 85 cm in women were considered to indicate abdominal obesity in the Chinese population and non-Hispanic Asian population in NHANES. The corresponding criteria in Hispanic and non-Hispanic White and Black were ≥ 102 cm in men and ≥ 88 cm in women. Abbreviations: DM=diabetes mellitus, BMI=body mass index, SBP=systolic blood pressure, DBP=diastolic blood pressure, HDLC=high-density lipoprotein cholesterol, LDLC=low-density lipoprotein cholesterol.

Supplementary Table 4 Apparent performance and internal validation of the prediabetes prediction model

Performance Statistics	Internal validation ^a				
	Apparent performance in original sample	Bootstrap performance in bootstrap sample	Test performance in original sample	Average optimism	Optimism corrected
C statistic (95% CI)	0.6998 (0.6933, 0.7063)	0.7002	0.6996	0.0006	0.6992
Calibration	Slope: 1.0002 Intercept: 0.000378	Slope: 1.0002 Intercept: 0.000345	Slope: 0.9978 Intercept: -0.002947	0.0024	Slope: 0.9978

^aPrediction model was internally validated in the CNDMDS dataset. CI=Confidence interval. The estimates (95% CI) of the apparent performance of the prediction model in the CNDMDS sample were 1.0002 (0.9615, 1.039) for the calibration slope and 0.000378 (-0.0620, 0.0627) for the slope.

Supplementary Table 5 Recalibration of the prediction model for prediabetes in the CNDMDS

Variable	OR 95% CI in original model	P value in original model	β coefficients in original model	Shrinkage factor	Shrunken predictor coefficients and recalculated intercept*
Intercept	-	<0.0001	-6.1358	-	-6.1262 ¹
Age, years	1.028(1.025-1.030)	<0.0001	0.0274	0.9978	0.0273
Middle school or higher education	0.816(0.766-0.870)	<0.0001	-0.2032	0.9978	-0.2028
DM family history	1.364(1.261-1.477)	<0.0001	0.3108	0.9978	0.3101
Waist circumference \geq normal upper limit	1.216(1.134-1.305)	<0.0001	0.1959	0.9978	0.1955
BMI	1.079(1.069-1.089)	<0.0001	0.0761	0.9978	0.0759
SBP	1.012(1.010-1.013)	<0.0001	0.0117	0.9978	0.0117

¹Predictor coefficients were shrunken by multiplying the shrinkage factor (average calibration slope of each bootstrapped sample) with the original coefficients of predictors in the original model. To maintain the overall apparent calibration, the intercept was recalculated by subtracting multiplications of shrunken coefficients with predictor values from the original linear prediction. The recalculated intercept was the mean difference between the original linear prediction and the summed multiplications of shrunken coefficients and predictor values.

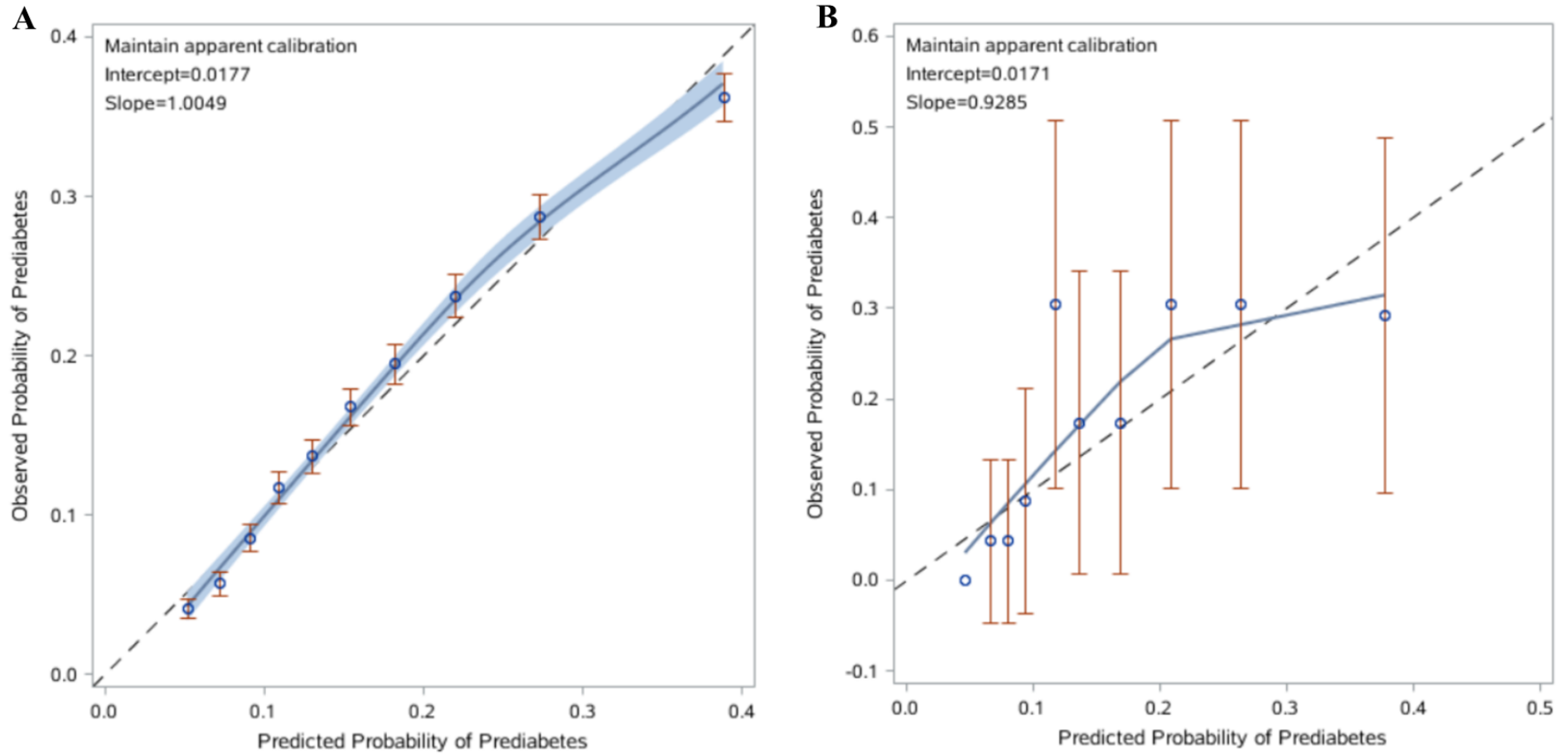
Supplementary Table 6 External validation performance of the prediabetes prediction model in diverse populations in NHANES

Performance Statistics	External validation in diverse populations ^b				
	Mexican American	Other Hispanic	None-Hispanic White	None-Hispanic Black	None-Hispanic Asian
C statistic (95% CI)	0.7613 (0.6725, 0.8502)	0.7776 (0.7022, 0.8530)	0.6926 (0.6449, 0.7404)	0.7441 (0.6828, 0.8053)	0.7090 (0.6324, 0.7857)
Calibration	Slope: 1.4497 Intercept: 0.470684	Slope: 1.3861 Intercept: 0.300176	Slope: 0.8646 Intercept: -0.311992	Slope: 1.2191 Intercept: -0.159560	Slope: 0.9238 Intercept: 0.000376

^bExternal validation was performed in the NHANES 2012 dataset. CI=confidence interval. At external validation, the calibration slope in NHANES 2012 reflects the combined effect of overfitting on the development data (CNDMDS) and true differences in the effects of predictors.

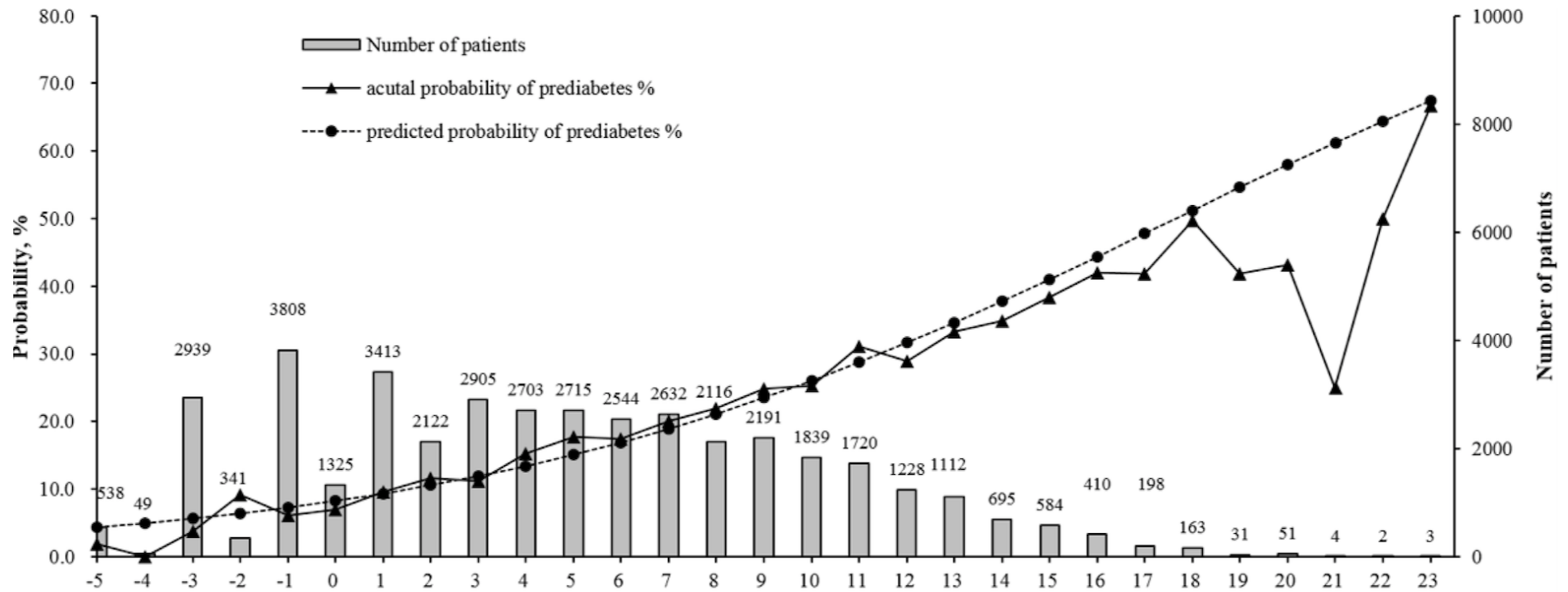
Supplementary Table 7 Total points and risk of prediabetes in the CNDMDS

Total point	Number of patients	Actual probability of prediabetes %	Predicted probability of prediabetes %
-5	538	1.859	4.335
-4	49	0	4.939
-3	2939	3.709	5.621
-2	341	9.091	6.391
-1	3808	6.119	7.259
0	1325	6.943	8.234
1	3413	9.581	9.327
2	2122	11.593	10.549
3	2905	11.256	11.909
4	2703	15.279	13.419
5	2715	17.68	15.087
6	2544	17.414	16.922
7	2632	20.023	18.93
8	2116	21.928	21.117
9	2191	24.783	23.482
10	1839	25.285	26.025
11	1720	31.047	28.74
12	1228	28.909	31.618
13	1112	33.273	34.643
14	695	34.82	37.798
15	584	38.356	41.059
16	410	41.951	44.401
17	198	41.919	47.795
18	163	49.693	51.209
19	31	41.935	54.612
20	51	43.137	57.972
21	4	25	61.26
22	2	50	64.448
23	3	66.667	67.513



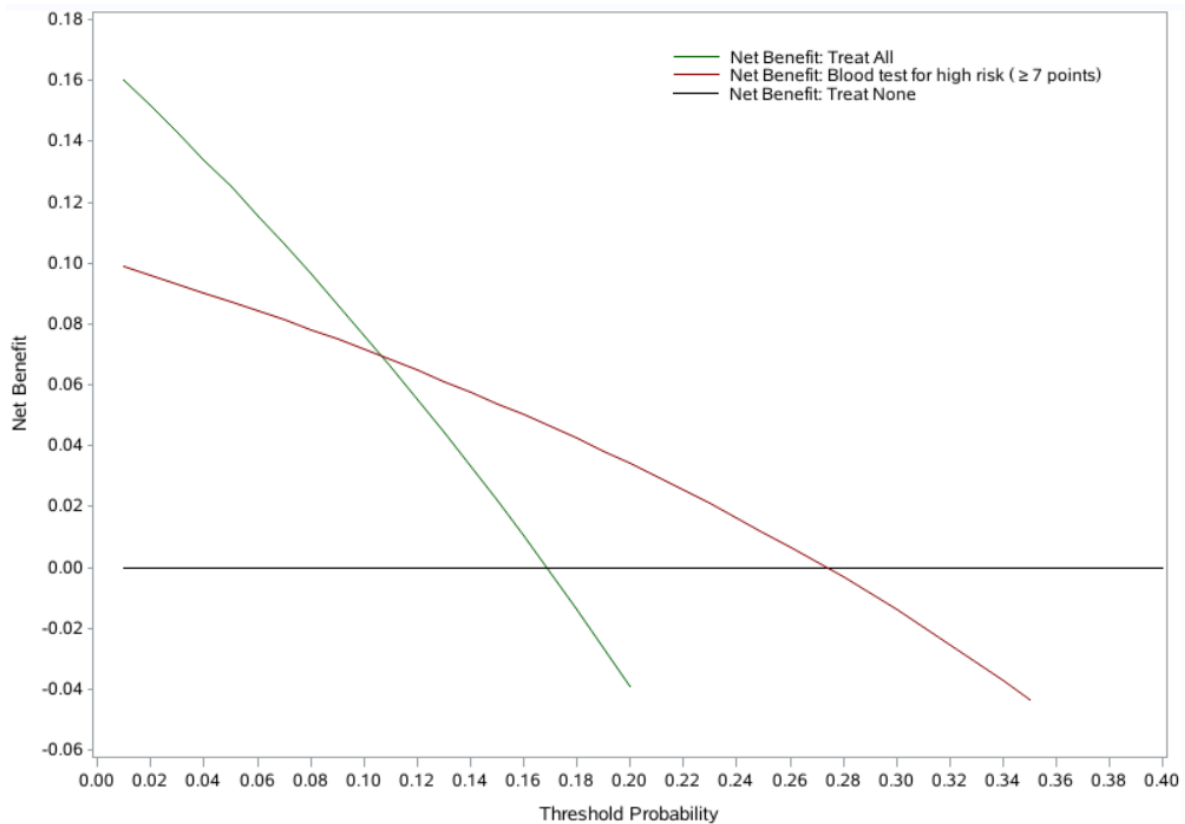
Supplementary Figure 1 Calibration plot of the prediction model with the recalibrated intercept and shrunken predictors: A: internal validation in CNDMDS, using prediction model with adjusted intercept -6.1262 and shrunken coefficient by 0.9978; B: external validation in the

Asian population in NHANES using a prediction model with an adjusted intercept of -6.1262 and a shrunken coefficient of 0.9978. The Hosmer and Lemeshow test indicated adequate calibration of our prediction model in the Asian population in NHANES (Chi-square=12.1014, P=0.1467)



Supplementary Figure 2 Observed and predicted probability of prediabetes in the CNDMDS by total risk points and distribution of patient

numbers. The observed and predicted probabilities of prediabetes over different risk points are demonstrated by solid and dot-dashed lines, respectively. The observed and predicted probabilities both increased with the rise of the risk points. The distribution of patient numbers over each risk point is presented by gray columns.



Supplementary Figure 3 Net benefit analysis. A better net benefit was produced for individuals with a total score ≥ 7 points to undergo a blood test for prediabetes detection, particularly in populations with prediabetes risk ranging between 0.12 and 0.28. The graph shows the expected net benefit per patient. The black line assumes that all participants do not undergo blood testing, the green line assumes testing all, and the red line indicates testing individuals with a total score ≥ 7 based on the prediction model.