Supplementary Materials.

Supplementary Table 1. Definitions of cardiovascular diseases

	ICD-9-CM	ICD-10-CM
Myocardial infarction	410.*	I21.*, I22.*, I23.3, I24.0, I24.9, I25.9, I51.3,
Coronary artery disease	411.*, 413.*, 414.*	I20.*, I23.7, I24.*, I25.*, T82.85,
Ischemic stroke	433.01, 433.11, 433.21, 433.31, 434.81, 433.91, 434.11, 434.91, 436.*	G46.*, I63.*, I67.85, I69.30, I77.89, P91.0, Z86.73
Hemorrhagic stroke	430.*, 431.*, 432.*	I60.*, I61.*, I62.*, I63.89, I67.1, P52.*, P54.8, S06.2*, S06.4*, S06.5*
Atrial fibrillation, heart failure	427.31, 428.*	I27.29, I48.0, I48.1, I48.2, I48.91, I50.*, I51.9, Z86.79
Other cardiovascular disease	412.* (MI), 346.6* (cerebrovascular), 433.*(cerebrovascular), 434.*(cerebrovascular), 435.* (cerebrovascular), 437.*(cerebrovascular), 438.*(cerebrovascular), 440.* (PAD), V45.81 (CAD), or V45.82 (CAD)	G43.6* (cerebrovascular), G45.* (cerebrovascular), G46.* (cerebrovascular), G81.94 (cerebrovascular), G83.9 (cerebrovascular), G93.89 (cerebrovascular), G93.9 (cerebrovascular), G95.19 (cerebrovascular), I25.1 (CAD), I25.2 (MI), I63.*(cerebrovascular), I65.* (cerebrovascular), I66.* (cerebrovascular), I67.* (cerebrovascular) , I68.* (cerebrovascular), I69.* (cerebrovascular), I70.* (PAD), I72.5 (cerebrovascular), I76 (vascular), I77.75 (cerebrovascular), I77.89 (cerebrovascular), I77.9 (cerebrovascular), I99.8 (vascular), M47.01 (cerebrovascular), P91.0 (cerebrovascular), Z98.61 (CAD)

Abbreviations: ICD-9-CM–International classification of diseases, 9th revision; ICD-10-CM–International classification of diseases, 10th revision

Supplementary Table 2. Hyperparameters searched for the machine learning models.

Model	Hyperparameter	Values
LR	C	0.01, 0.1, 1, 10
Lasso	C	0.001, 0.01,0.1, 1, 10
RF	n_estimators	100, 200, 500, 1000, 2000, 4000, 5000
	max_depth	3, 5, 7, 9
	min_samples_split	2
	max_features	0.5, 1
GBM	n_estimators	20, 30, 50, 100, 200, 300, 500, 1000, 2000, 3000, 4000
	max_depth	3, 4, 5, 7
	subsample	0.75, 1
	max_features	0.5, 1
XGB	n_estimators	20, 30, 50, 100, 200, 300, 500, 1000
	max_depth	3, 5, 7
	learning_rate	0.3

	reg_alpha	0,1,10
	reg_lambda	0,1,10

Abbreviations: LR–logistic regression; Lasso–logistic regression with a lasso penalty; RF–random forest; GBM–gradient boosting machine; XGBoost–extreme gradient boosting

Supplementary Table 3. Missingness of variables.

	Missing, N (%)
Race	5085 (15.8)
HDL cholesterol	5966 (18.5)
LDL cholesterol	5887 (18.3)
Total cholesterol	5934 (18.4)
Systolic BP	956 (3)
Diastolic BP	960 (3)
Current smoker	588 (1.8)
Height	14517 (45.1)
Weight	2884 (9)
Median household income	2290 (7.1)
Education level	2287 (7.1)

Abbreviations: CVD – cardiovascular disease, BP – blood pressure; LDL – low density lipoprotein; HDL – high density lipoprotein

Supplementary List 1. Complete list of all variables used in the final models.

Clinical Variables
Age
Female
Race
Weight
Height
HDL cholesterol
Total cholesterol
LDL cholesterol
Systolic BP
Diastolic BP
Current smoker

Healthcare utilization variables
N primary care visits
N other service visits
N specialty care visits

Socioeconomic variables
Median household income
Percent with up to an associate's degree
Percent with up to a high school graduation
Percent with up to some college
Percent with up to a bachelor's degree
Percent with up to a 9th grade education
Percent with up to 12th grade education

Family history variables
Family history of emphysema

Lab test variables
N normal lab test results
N abnormal lab test results
Aspartate aminotransferase lab: normal
Basophils lab: normal
Bilirubin lab: normal
Bilirubin lab: abnormal
Urea nitrogen lab: abnormal
Calcium lab: normal
Eosinophils lab: normal
Eosinophils lab: abnormal
Fibrinogen lab: abnormal
Glomerular filtration rate lab: abnormal
Glucose lab: abnormal
Potassium lab: normal
Lymphocytes lab: abnormal
Monocytes lab: abnormal
Prostate specific Ag lab: normal
Thyroxine lab: abnormal
Albumin/Creatinine lab: abnormal
Albumin lab: abnormal
Protein lab: abnormal

Medication variables

On antihypertensive medications
Total Medications Prescribed
GPI4: Penicillin Combinations
GPI4: Aminopenicillins
GPI4: Cephalosporins - 3rd Generation
GPI4: Erythromycins
GPI4: Aminoglycosides
GPI4: CMV Agents
GPI4: Leprostatics
GPI4: Beta Blockers Cardio-Selective
GPI4: ACE Inhibitors
GPI4: Angiotensin II Receptor Antagonists
GPI4: Loop Diuretics
GPI4: Thiazides and Thiazide-Like Diuretics
GPI4: HMG CoA Reductase Inhibitors
GPI4: Antihyperlipidemics - Combinations
GPI4: Antihistamines - Non-Sedating
GPI4: Nasal Antiallergy
GPI4: Bulk Laxatives
GPI4: Surfactant Laxatives
GPI4: Antidiarrheal Combinations
GPI4: H-2 Antagonists
GPI4: Antipsychotics - Misc.
GPI4: Antidementia Agents
GPI4: Salicylates
GPI4: Water Soluble Vitamins
GPI4: B-Complex w/ C
GPI4: B-Complex w/ Folic Acid
GPI4: Vitamins w/ Lipotropics
GPI4: Water Soluble Vitamins
GPI4: Infant Care Products
GPI4: Iron
GPI4: Beta-blockers - Ophthalmic
GPI4: Ophthalmic Immunomodulators
GPI4: Emollient/Keratolytic Agents
GPI4: Emollients
GPI4: Bulk Chemicals - C's
GPI4: Central Muscle Relaxants
GPI4: Cough/Cold/Allergy Combinations
GPI4: Immunosuppressive Agents

Comorbidity variables
History of Type 2 Diabetes
CCS: Nonspecific chest pain
CCS: Other and ill-defined heart disease
CCS: Cancer of head and neck
CCS: Aortic; peripheral; and visceral artery aneurysms
CCS: Pancreatic disorders (not diabetes)
CCS: Chronic kidney disease
CCS: Chronic ulcer of skin
CCS: Septicemia (except in labor)
CCS: Other non-traumatic joint disorders
CCS: Other connective tissue disease
CCS: Nervous system congenital anomalies
CCS: Other injuries and conditions due to external causes

CCS: Cancer of kidney and renal pelvis
CCS: Diabetes mellitus without complication
CCS: HIV infection
CCS: Diabetes mellitus with complications
CCS: Other endocrine disorders
CCS: Deficiency and other anemia
CCS: Diseases of white blood cells
CCS: Other hematologic conditions
CCS: Anxiety disorders
CCS: Delirium, dementia, and amnestic and other cognitive disorders
CCS: Alcohol-related disorders
CCS: Parkinson's disease
CCS: Other infections; including parasitic
CCS: Epilepsy; convulsions
CCS: Other nervous system disorders
CCS: Heart valve disorders
CCS: Peri-; endo-; and myocarditis; cardiomyopathy (except that caused by tuberculosis or sexually transmitted disease)
CCS: Essential hypertension
CCS: Hypertension with complications and secondary hypertension

Supplementary Table 4. Performance metrics of XGBoost model in the test set of the full CVD cohort.

Percentage of patients	Sensitivity (%)	Specificity (%)	PPV (%)	NNE	Youden Index	F1 Score
1%	3	99	36	3	0.02	0.05
2%	5	98	32	3	0.04	0.09
5%	11	96	28	4	0.07	0.16
10%	20	91	25	4	0.12	0.23
25%	46	78	23	4	0.23	0.30
50%	74	53	18	5	0.27	0.29
75%	94	28	16	6	0.22	0.27
90%	98	11	14	7	0.09	0.24
95%	99	6	13	8	0.04	0.23
98%	99	2	13	8	0.01	0.22
99%	100	1	12	8	0.01	0.22

Supplementary Table 5. Performance metrics of XGBoost model in the test set of the ASCVD cohort.

Percentage of patients	Sensitivity (%)	Specificity (%)	PPV (%)	NNE	Youden Index	F1 Score
1%	2	99	30	3	0.02	0.05
2%	5	98	33	3	0.04	0.09
5%	11	96	27	4	0.07	0.16
10%	23	92	27	4	0.15	0.25
25%	49	78	23	4	0.27	0.31
50%	75	53	18	6	0.29	0.29
75%	94	28	15	7	0.21	0.26
90%	98	11	13	8	0.09	0.23
95%	99	6	13	8	0.05	0.22
98%	100	2	12	8	0.02	0.22
99%	100	1	12	8	0.01	0.22