SUPPLEMENTARY DATA

Supplementary Table 1. ACCORD Variables Available for Risk Prediction Model. **Bolded** candidate variables had <10% missing data and were considered in risk prediction models.

Variable Domain	Individual Variables
Demographic/Anthropometric	**Age**, **sex**, **race**, **body mass index, systolic blood pressure, diastolic blood pressure**, **heart rate**, **randomized treatment arm**, **body mass index**, **waist circumference**
Clinical/Medical History	**Systolic blood pressure, diastolic blood pressure**, **heart rate**, **History of myocardial infarction, stroke, coronary artery bypass graft surgery, cardiovascular disease, percutaneous coronary intervention/atherectomy, albuminuria**, **low ankle-brachial index**, **current smoker**, **level of education, years of diabetes mellitus, hypertension, hyperlipidemia**, **family history of heart failure**, hypoglycemic event requiring medical assistance, hypoglycemic event requiring any assistance, **neuropathy**, **vibration sensation loss**, **pressure sensation loss**, Mini–Mental State Examination total score, Denver Developmental Screening Test total score, Stroop score, Rey Auditory Verbal Learning Test score, **former cigarette smoker, lives with one or more people, foot ulcer requiring antibiotics**, **number of alcoholic drinks per week, Health Utilities Index Mark3 score, Health Utilities Index Mark2 score**, physical health survey, mental health survey, diet summary score, total exercise expenditure calories/week, total moderate-intensive exercise expenditure calories/week, documented blood glucose <50mg/dl, serum creatinine doubling, right leg pre-tibial edema, left leg pre-tibial edema, amputation of lower extremity, photocoagulation, cataract extraction, severe vision loss, **Michigan Neuropathy Screening Instrument score**, **visual acuity score R eye**, **visual acuity score L eye**, **depression**, **eye disease**, full insurance benefits, partial insurance benefits, **foot appearance**, **ankle reflexes**, **neuropathy screening instrument score**, **feeling score**, nutrition education
Laboratory	**Potassium, serum creatinine, glycated hemoglobin, total cholesterol, triglycerides, very low-density lipoprotein, low-density lipoprotein, HDL-C, fasting plasma glucose, alanine aminotransferase, creatine kinase, estimated glomerular filtration rate, urinary albumin, urinary creatinine**
Baseline Medications	**Loop diuretic, thiazide diuretic, potassium-sparing diuretic, angiotensin II receptor blocker, angiotensin-converting enzyme inhibitor, dihydropyridine calcium channel blockers, non-dihydropyridine calcium channel blockers, alpha-blockers, beta-blockers**, **vasodilators**, **reserpine**, **digitalis**, **anti-arrhythmic therapies**, **nitrates**, **sulfonylureas**, **biguanides**, **meglitinides**, **alpha-glucosidase inhibitors**, **NPH insulin**, **thiazolidinediones**, **regular insulin**, **Lispro/Aspart insulin**, **premixed insulin**, **bile-acid sequestrants**, **statins**, **fibrates**, **cholesterol absorption inhibitors**, **niacin**, **oral anticoagulants**, **nonsteroidal anti-inflammatory drugs, cyclooxygenase-2 inhibitors, aspirin,**

	thyroid agent, **progestins**, **estrogens**, **oral asthma drugs**, **any antidepressant**, **drugs for osteoporosis**, **vitamins**, **over-the-counter medication**, **herbal medication**
Neuro Magnetic Resonance Imaging	Intracranial volume, basal ganglia normal, basal ganglia abnormal, gray mater normal, white matter normal, gray matter abnormal, white matter abnormal, brain volume normal, brain volume abnormal, gray matter total, white matter total
Electrocardiographic	ST-segment depression anterolateral, posterior, anterior, T-wave inversion, AV conduction defect, **PR interval, QRS duration, QTc interval, QRS-axis**, **P-axis, T-axis**, **R-wave amplitude (lead aVL), S-wave amplitude (lead V3), left ventricular hypertrophy (Cornell Voltage), atrial fibrillation**

SUPPLEMENTARY DATA

Supplementary Table 2. Identified variables in each of the three variable selection methods: random survival forest (RSF), stepwise forward selection, and stepwise backwards selection.

Random Survival Forest (RSF)	**Stepwise Forward Selection**	**Stepwise Backwards Selection**
Age	Age	Age
T-wave axis*	T-wave axis*	T-wave axis*
QTc duration*	Hemoglobin A1c*	Hemoglobin A1c*
Diastolic blood pressure	Serum creatinine	Urine albumin
Hemoglobin A1c*	Diastolic blood pressure	Serum creatinine
Serum creatinine	QTc duration*	Serum potassium
HDL-C	History of MI	QRS duration
QRS duration	PR duration*	Fasting plasma glucose
Body mass index	Urine creatinine	History of MI
History of MI	HDL-C	History of CABG
Fasting plasma glucose	Body mass index	PR duration*
History of CABG	QRS duration	Urine creatinine
Systolic blood pressure		HDL-C
		Diastolic blood pressure

* Indicates the variable was removed from the relationship modeling step as the variable was not a routine clinical marker, clinically similar to other variables, or not available in the validation dataset.

SUPPLEMENTARY DATA

Supplementary Table 3. Cox proportional hazard parameter estimates for the estimated 5-year risk of HF.

Variable	β-estimate	Hazard Ratio (95% Confidence Interval))	P-value
Age	0.039	1.04 (1.02-1.06)	<0.001
Fasting plasma glucose	0.002	1.00 (1.00-1.01)	0.03
Body mass index	0.057	1.06 (1.04-1.08)	<0.001
Systolic blood pressure	0.015	1.02 (1.01-1.02)	<0.001
Diastolic blood pressure	-0.025	0.98 (0.96-0.99)	<0.001
Serum creatinine	0.710	2.03 (1.31-3.16)	0.002
HDL-C	-0.023	0.98 (0.97-0.99)	<0.001
QRS duration (≥120 msec)	0.583	1.79 (1.31-2.45)	<0.001
History of MI	0.600	1.82 (1.37-2.41)	<0.001
History of CABG	0.673	1.96 (1.46-2.64)	<0.001

Baseline survival: 0.97358

SUPPLEMENTARY DATA

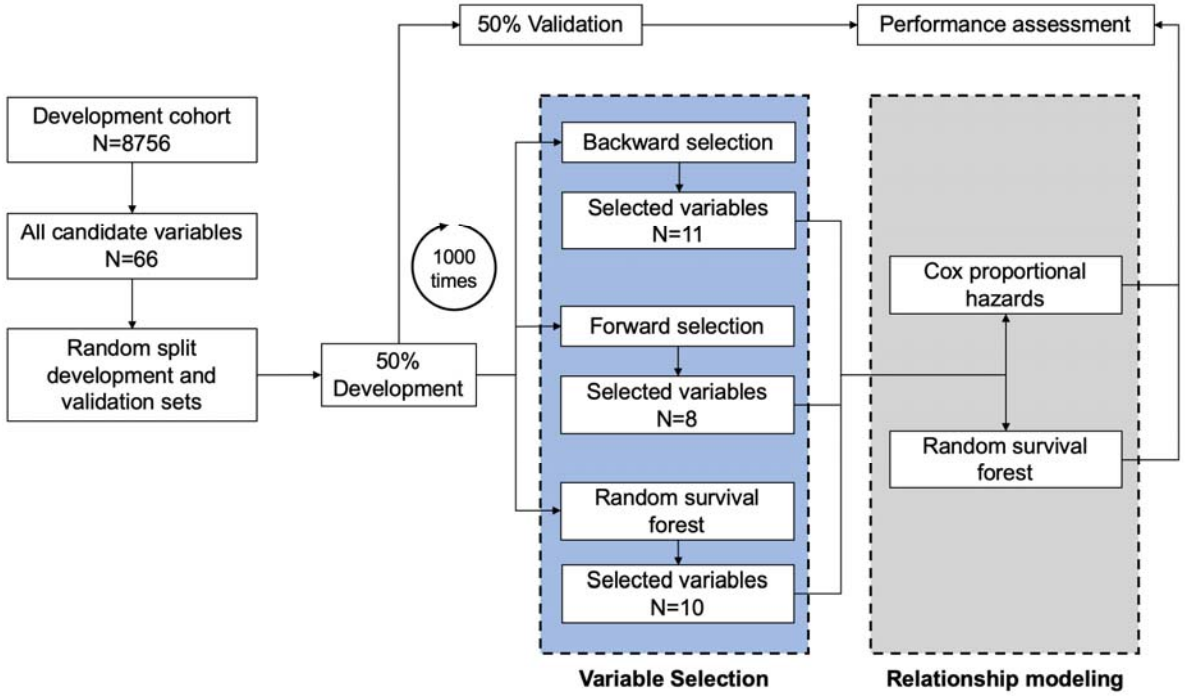
Supplementary Table 4. Baseline characteristics for both the ACCORD and ALLHAT datasets.

	ACCORD (n=8,756)	**ALLHAT (n=10,819)**	***P*-value**
Female, %	3370 (38.5)	5306 (49.0)	<0.01
Age (years), %	62.7 (6.6)	67.0 (7.3)	<0.01
Race, %			<0.01
Black	1622 (18.5)	4109 (38.0)	
Other	1645 (18.8)	580 (5.4)	
White	5489 (62.7)	6130 (56.7)	
Systolic blood pressure (mmHg), sd	136.4 (16.9)	144.7 (13.5)	<0.01
Diastolic blood pressure (mmHg), sd	75.0 (10.5)	82.0 (9.9)	<0.01
Body mass index (kg/m^2), sd	32.1 (5.4)	31.0 (6.0)	<0.01
QRS duration (>120 ms), %	636 (7.3)	806 (7.4)	0.64
LV hypertrophy[1], %	332 (3.8)	983 (9.1)	<0.01
Fasting plasma glucose (md/dL), sd	175.2 (55.9)	172.4 (74.5)	0.18
Potassium (md/dL), sd	4.5 (0.5)	4.4 (0.7)	0.39
Serum creatinine (md/dL), sd	0.9 (0.2)	1.0 (0.3)	0.31
Estimated glomerular filtration rate (mL/min), sd	91.5 (27.3)	80.7 (21.4)	<0.01
Total cholesterol (mg/dL), sd	183.6 (41.7)	215.4 (45.5)	<0.01
Triglycerides (md/dL), sd	190.6 (149.8)	191.1 (153.0)	0.87
Low-density lipoprotein cholesterol (md/dL), sd	105.1 (33.9)	134.1 (38.8)	<0.01
HDL-C (md/dL), sd	41.9 (11.5)	44.8 (12.4)	<0.01
History of myocardial infarction, %	1237 (14.1)	2063 (19.1)	<0.01
History of stroke, %	502 (5.7)	728 (6.7)	0.29
History of coronary artery bypass graft surgery, %	918 (10.5)	1141 (10.5)	0.91
Current cigarette smoker, %	1078 (12.3)	1342 (12.4)	0.86

[1]Determined by Cornell Voltage criteria

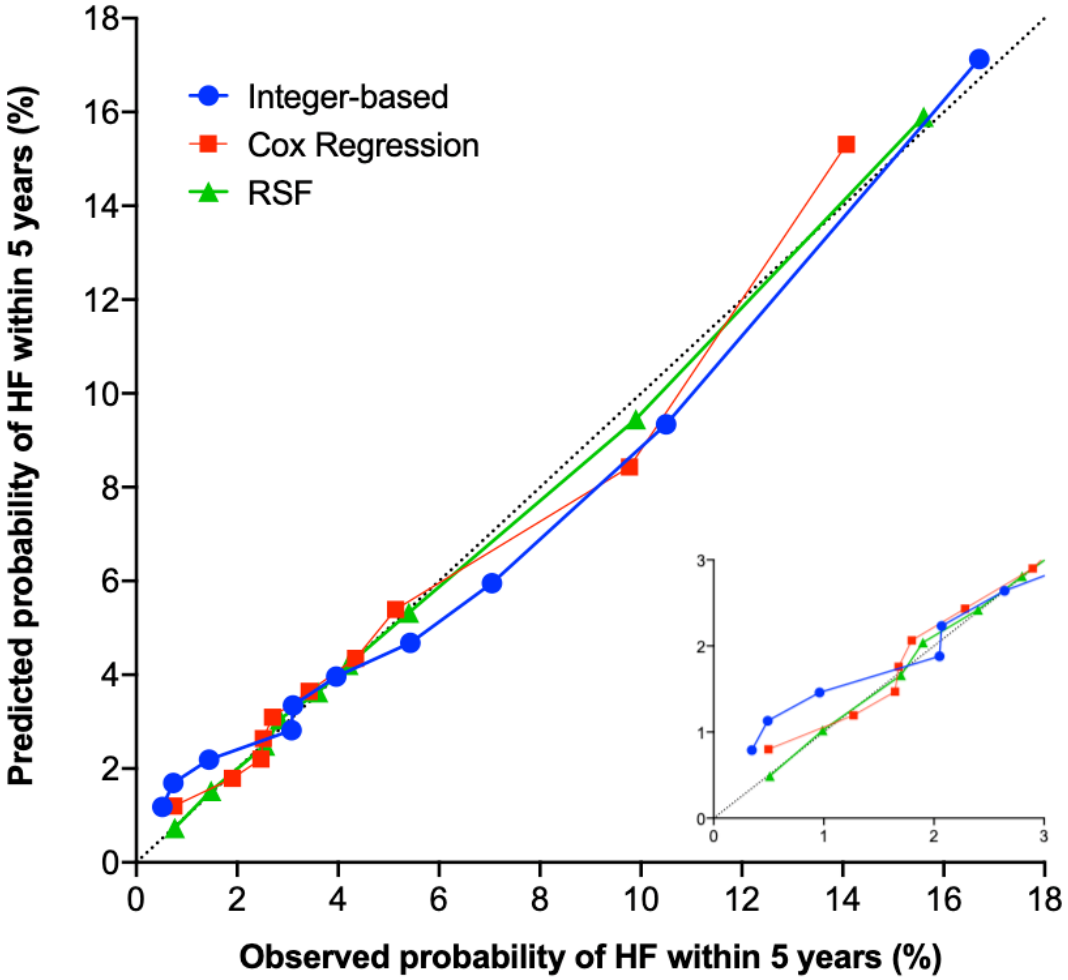
SUPPLEMENTARY DATA

Supplementary Figure 1. The analysis flow for developing and validating models.



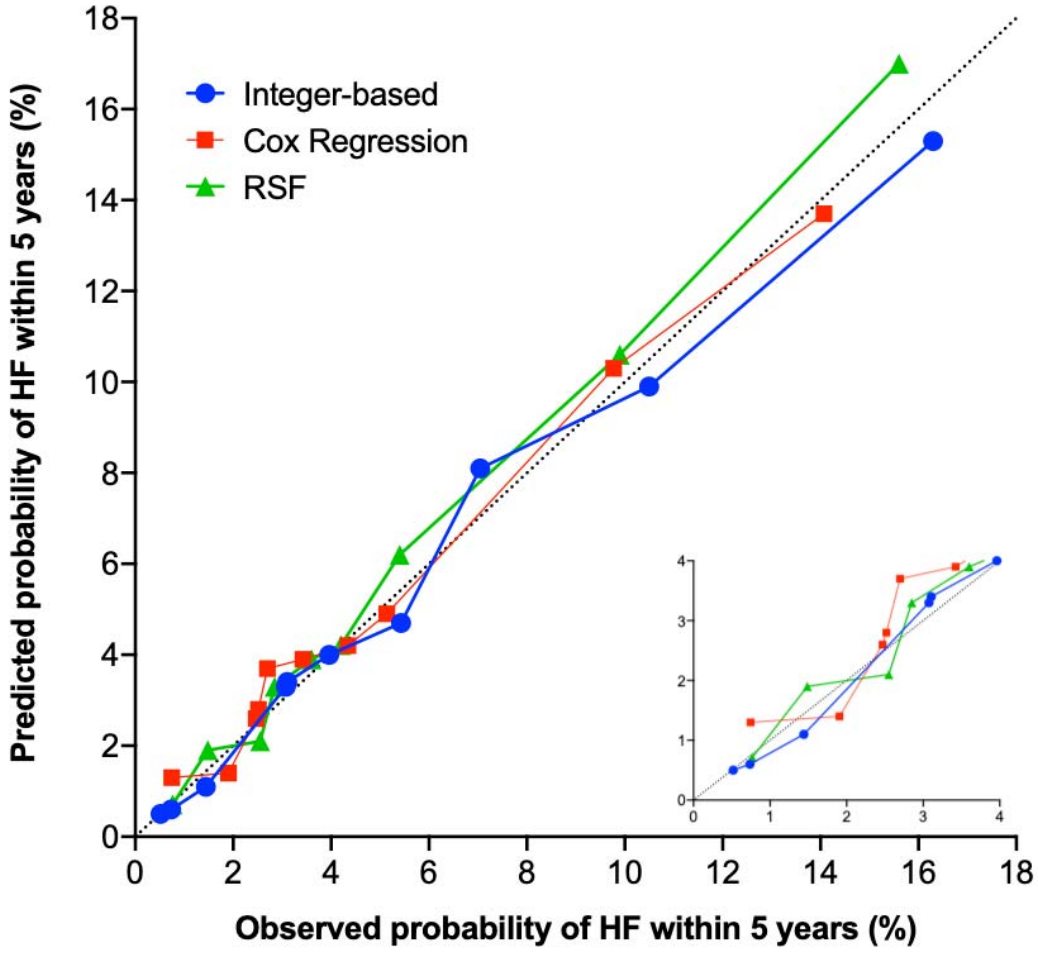
SUPPLEMENTARY DATA

Supplementary Figure 2. Calibration of the WATCH-DM risk score for the integer-based, Cox regression-based, and RSF-based models in the internally validated dataset (ACCORD). Predicted vs observed 5-year incidence of heart failure based on deciles of predicted risk. Calibration was acceptable (Hosmer-Lemeshow statistics p≥0.20 for all models).



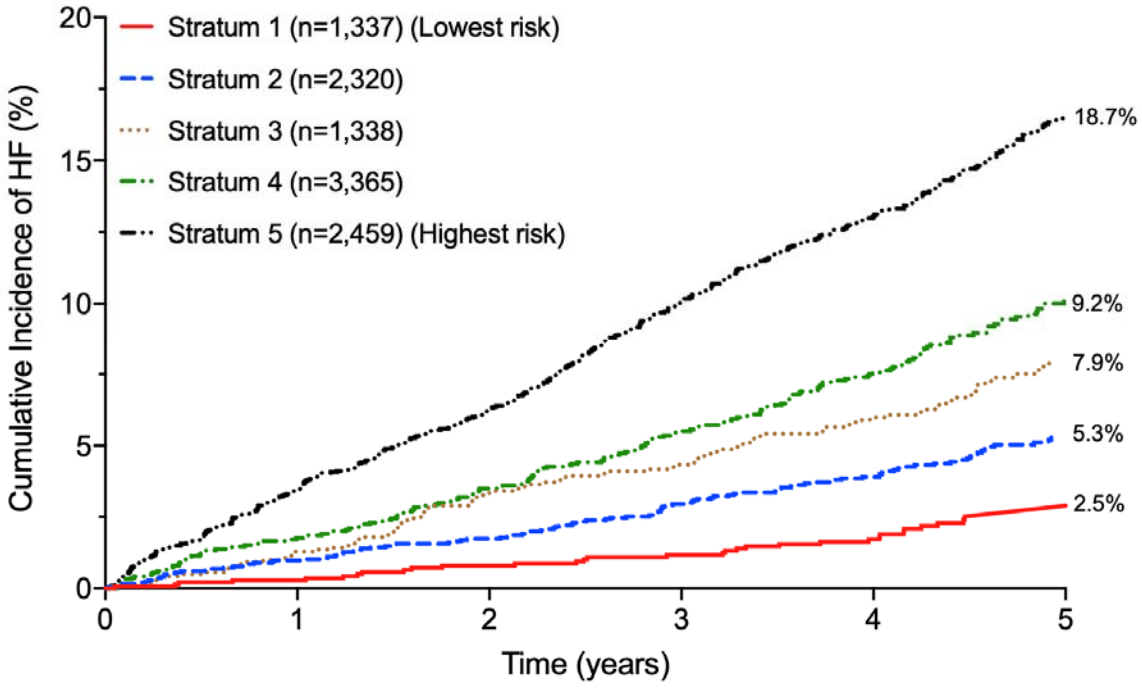
SUPPLEMENTARY DATA

Supplementary Figure 3. Calibration of the WATCH-DM risk score for the integer-based, Cox regression-based, and RSF-based models in the externally validated dataset (ALLHAT). Predicted vs observed 5-year incidence of heart failure based on deciles of predicted risk. Calibration was acceptable (Hosmer-Lemeshow statistics p≥0.10 for all models).



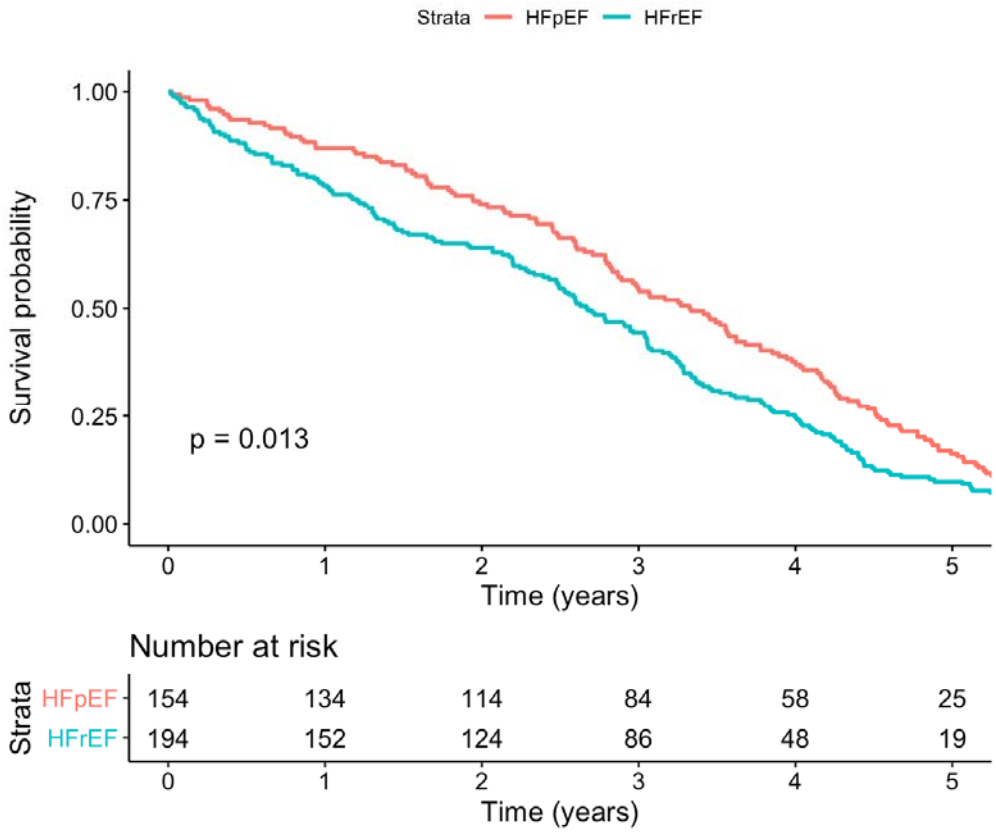
SUPPLEMENTARY DATA

Supplementary Figure 4. Cumulative incidence of HF across quintiles of WATCH-DM: Q1 (≤7); Q2 (8-9); Q3 (10); Q4 (11-13); Q5 (≥14) in the validation dataset, ALLHAT.



SUPPLEMENTARY DATA

Supplementary Figure 5. Kaplan-Meier curves for HF events by phenotype in the ALLHAT (external validation) cohort. The *p*-value was calculated using the *logrank* test.



SUPPLEMENTARY DATA

Supplementary Figure 6. Cumulative incidence curves for incident HF in each of the 5 risk strata by HF phenotype in the external validation (ALLHAT) dataset.