Journal of the American Heart Association

ORIGINAL RESEARCH

Validation of the WATCH-DM and TRS-HF$_{DM}$ Risk Scores to Predict the Risk of Incident Hospitalization for Heart Failure Among Adults With Type 2 Diabetes: A Multicohort Analysis

Matthew W. Segar, MD, MS*; Kershaw V. Patel, MD, MSCS*; Anne S. Hellkamp, MS; Muthiah Vaduganathan, MD, MPH; Yuliya Lokhnygina, PhD; Jennifer B. Green, MD; Siu-Hin Wan, MD; Ahmed A. Kolkailah, MD, MSc; Rury R. Holman, MBChB; Eric D. Peterson, MD, MPH; Vaishnavi Kannan, MS; Duwayne L. Willett, MD, MS; Darren K. McGuire, MD, MHSc; Ambarish Pandey, MD, MSCS

BACKGROUND: The WATCH-DM (weight [body mass index], age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control [fasting plasma glucose], ECG QRS duration, myocardial infarction, and coronary artery bypass grafting) and TRS-HF$_{DM}$ (Thrombolysis in Myocardial Infarction [TIMI] risk score for heart failure in diabetes) risk scores were developed to predict risk of heart failure (HF) among individuals with type 2 diabetes. WATCH-DM was developed to predict incident HF, whereas TRS-HF$_{DM}$ predicts HF hospitalization among patients with and without a prior HF history. We evaluated the model performance of both scores to predict incident HF events among patients with type 2 diabetes and no history of HF hospitalization across different cohorts and clinical settings with varying baseline risk.

METHODS AND RESULTS: Incident HF risk was estimated by the integer-based WATCH-DM and TRS-HF$_{DM}$ scores in participants with type 2 diabetes free of baseline HF from 2 randomized clinical trials (TECOS [Trial Evaluating Cardiovascular Outcomes With Sitagliptin], N=12 028; and Look AHEAD [Look Action for Health in Diabetes] trial, N=4867). The integer-based WATCH-DM score was also validated in electronic health record data from a single large health care system (N=7475). Model discrimination was assessed by the Harrell concordance index and calibration by the Greenwood-Nam-D'Agostino statistic. HF incidence rate was 7.5, 3.9, and 4.1 per 1000 person-years in the TECOS, Look AHEAD trial, and electronic health record cohorts, respectively. Integer-based WATCH-DM and TRS-HF$_{DM}$ scores had similar discrimination and calibration for predicting 5-year HF risk in the Look AHEAD trial cohort (concordance indexes=0.70; Greenwood-Nam-D'Agostino P>0.30 for both). Both scores had lower discrimination and underpredicted HF risk in the TECOS cohort (concordance indexes=0.65 and 0.66, respectively; Greenwood-Nam-D'Agostino P<0.001 for both). In the electronic health record cohort, the integer-based WATCH-DM score demonstrated a concordance index of 0.73 with adequate calibration (Greenwood-Nam-D'Agostino P=0.96). TRS-HF$_{DM}$ score could not be validated in the electronic health record because of unavailability of data on urine albumin/creatinine ratio in most patients in the contemporary clinical practice.

CONCLUSIONS: The WATCH-DM and TRS-HF$_{DM}$ risk scores can discriminate risk of HF among intermediate-risk populations with type 2 diabetes.

Key Words: diabetes ■ heart failure ■ risk prediction ■ risk score

Correspondence to: Ambarish Pandey, MD MSCS, Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390. Email: ambarish.pandey@utsouthwestern.edu

*M. W. Segar and K. V. Patel contributed equally as co-first authors.

Supplemental Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.024094

For Sources of Funding and Disclosures, see page 10.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- This study demonstrates that the WATCH-DM (weight [body mass index], age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control [fasting plasma glucose], ECG QRS duration, myocardial infarction, and coronary artery bypass grafting) and TRS-HF_{DM} (Thrombolysis in Myocardial Infarction [TIMI] risk score for heart failure in diabetes) risk scores can discriminate risk of heart failure (HF) among low- and intermediate-risk populations with type 2 diabetes.
- Among high-risk cohorts, neither risk score was well calibrated, and they tended to underestimate HF risk.
- The WATCH-DM risk score can be calculated from data routinely collected in the electronic health record.

What Are the Clinical Implications?

- The WATCH-DM and TRS-HF_{DM} risk scores were able to stratify HF risk among adults with type 2 diabetes and differing degrees of baseline cardiovascular disease risk.
- Future studies are needed to evaluate whether diabetes-specific risk scores can improve use of effective preventive interventions, such as sodium-glucose cotransporter-2 inhibitors, to lower HF risk.

Nonstandard Abbreviations and Acronyms

FPG	fasting plasma glucose
Look AHEAD	Look Action for Health in Diabetes
SGLT2i	sodium-glucose cotransporter-2 inhibitors
T2D	type 2 diabetes
TECOS	Trial Evaluating Cardiovascular Outcomes With Sitagliptin
TRS-HFDM	Thrombolysis in Myocardial Infarction (TIMI) risk score for heart failure in diabetes
UACR	urine albumin/creatinine ratio
WATCH-DM	weight (body mass index), age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control (fasting plasma glucose), ECG QRS duration, myocardial infarction, and coronary artery bypass grafting

Type 2 diabetes (T2D) affects >30 million adults in the United States and is an independent risk factor for heart failure (HF).[1,2] Over the past 2 decades, there has been a shift in cardiovascular complications observed in T2D with greater hospitalizations for HF compared with atherosclerotic cardiovascular disease.[3,4] Moreover, individuals with both T2D and HF are subject to a higher risk of all-cause and cardiovascular death.[5] Even with adequate control of glycemic status and other cardiovascular risk factors, the increased risk of HF among individuals with T2D persists, highlighting the need for novel approaches to its prevention.[6,7] Recent therapeutic advances in pharmacotherapies, such as sodium-glucose cotransporter-2 inhibitors (SGLT2i), have shown to be beneficial in preventing HF among patients with T2D.[8] However, the uptake of these therapies in patients with T2D has been low.[9] Identification of individuals with T2D who are at the highest risk of developing HF is key for efficient and cost-effective allocation of preventive therapies. To this end, recent studies have focused on developing simple and accurate risk scores using clinical, laboratory, and electrocardiographic variables to predict risk of HF development over short-term follow-up.[10–13]

Among HF risk scores developed specifically for patients with T2D, the WATCH-DM (weight [body mass index], age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control [fasting plasma glucose {FPG}], ECG QRS duration, myocardial infarction [MI], and coronary artery bypass grafting [CABG]) and TRS-HF_{DM} (Thrombolysis in Myocardial Infarction [TIMI] risk score for heart failure in diabetes) risk scores have demonstrated good performance to predict short-term risk of HF among adults with T2D.[10,11] WATCH-DM was developed to predict 5-year incident HF risk using data from the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial (median follow-up, 4.9 years) and externally validated among participants with diabetes from the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) (median follow-up, 4.8 years) and pooled community cohorts (up to 5 years).[14-16] TRS-HF_{DM} was developed in the Saxagliptin Assessment of Vascular Outcomes Recorded in Patients With Diabetes Mellitus-Thrombolysis in Myocardial Infarction 53 (SAVOR-TIMI 53) trial (median follow up, 2.1 years) and externally validated in Dapagliflozin Effect on Cardiovascular Events-Thrombolysis in Myocardial Infarction 58 (DECLARE-TIMI 58) (median follow up, 4.2 years) and ACCORD trial cohorts to predict risk of HF hospitalization among patients with as well as without a history of HF.[11,17] HF risk stratification is more relevant in patients without a history of HF, especially among those with T2D, in whom there is considerable heterogeneity in risk for developing HF and SGLT2i have emerged as

important preventive therapies. In contrast, patients with established HF are at high risk for complications and warrant aggressive implementation of guideline-directed medical therapy, including SGLT2i. Both scores have demonstrated good performance in predicting HF risk among individuals without history of HF, highlighting their utility for risk stratification. However, performance of the WATCH-DM and TRS-HF_{DM} scores to predict incident HF risk across cohorts with different baseline risk and among contemporary patients outside the clinical trial setting is uncertain. Accordingly, in this study, we aimed to evaluate the performance of the WATCH-DM and TRS-HF_{DM} risk scores for incident HF risk prediction in 3 separate cohorts: participants with higher atherosclerotic cardiovascular disease burden from the TECOS (Trial Evaluating Cardiovascular Outcomes With Sitagliptin) and participants with lower atherosclerotic cardiovascular disease risk from the Look AHEAD (Look Action for Health in Diabetes) trial, and a real-world cohort of patients from the University of Texas (UT) Southwestern Medical Center.

METHODS

Our study data will not be made available to other researchers for purposes of reproducing the results because of institutional review board and clinical trial restrictions. We analyzed data from the TECOS and Look AHEAD clinical trials and from electronic health records (EHRs) at UT Southwestern Medical Center. In all cohorts comprising adults with T2D, participants with a history of HF were excluded. Data from the Look AHEAD trial were obtained from the National Institute of Diabetes and Digestive and Kidney Disease Repository. The TECOS database is located at the Duke Clinical Research Institute (Durham, NC). All trial participants provided written informed consent, and the studies were approved by the ethics committees for each participating trial site. Analyses were performed at UT Southwestern and the Duke Clinical Research Institute. The analyses of the UT Southwestern EHR data and the combined analyses presented herein were deemed exempt by the Institutional Review Board at the UT Southwestern Medical Center (Dallas, TX).

Study Populations

The details of the Look AHEAD trial have previously been published and are described (Data S1).[18] Briefly, the Look AHEAD trial was a randomized, multicenter, clinical trial involving 5145 participants with T2D randomized to either an intensive lifestyle intervention focused on weight loss by reduced caloric intake plus increased physical activity (intervention arm) or diabetes support and education alone (control arm) to determine the impact of an intensive lifestyle intervention on the development of cardiovascular disease. Among 5145 participants in the Look AHEAD trial cohort, 239 were not included in the National Institute of Diabetes and Digestive and Kidney Disease data set. We further excluded 39 participants with a history of HF. The final analysis cohort included 4867 participants.

The design and results of TECOS have previously been published and are detailed (Data S1).[19,20] Briefly, TECOS was a double-blind, multicenter, randomized trial of 14 671 participants evaluating the cardiovascular safety of the dipeptidyl peptidase 4 inhibitor sitagliptin versus placebo. Among 14 671 participants in the TECOS cohort, we excluded 2643 with a history of HF. The final analysis cohort included 12 028 participants.

The EHR cohort included patients at the UT Southwestern Medical Center, who were registered in the institution's diabetes registry as of December 31, 2014. Patients were included in the diabetes registry if they had an active problem of T2D listed in the EHR. Patients who were also listed in the institution's HF registry as of December 31, 2014, were excluded. Patients were similarly added to the HF registry if they had an active problem of HF listed or had a previous clinical encounter, either ambulatory or inpatient, for HF. The study cohort included 17 929 patients with a diagnosis of T2D and free of HF as of December 31, 2014. We excluded 4413 patients who did not have at least one recorded clinic visit before December 31, 2019, 3807 patients with laboratory data not recorded within 12 months of each other, and 2234 patients with >20% missingness of the WATCH-DM covariates. The final analysis cohort included 7475 patients.

Clinical Covariates

Clinical and laboratory values for the Look AHEAD trial and TECOS cohorts were obtained using a standardized protocol, as published previously and described (Data S1).[19,21] For the EHR cohort, the baseline data were abstracted from the health record from the closest visit before start of the follow-up period. A detailed description is provided (Data S1). Among participants from the Look AHEAD trial and EHR cohorts, missing data were imputed using random forest imputation.[22] In the TECOS cohort, missing data were imputed on the basis of the fully conditional specification method, taking into consideration the joint distribution of other variables.

Outcomes of Interest

The primary outcome of interest for the present analysis was incident HF hospitalization. In the Look AHEAD trial cohort, incident HF events were adjudicated by a committee of physicians blinded to randomization after reviewing medical records, including the medical

history, test results, and medication use, as reported previously.[18] The median follow-up in the Look AHEAD trial data available through the National Institute of Diabetes and Digestive and Kidney Disease was 9.6 (interquartile range [IQR], 8.9–10.3) years. Because the WATCH-DM and TRS-HF_{DM} risk scores were developed to predict short-term risk of HF (up to 5-year risk assessed in WATCH-DM and up to 4.8-year risk for TRS-HF_{DM} across different cohorts),[10,11] the Look AHEAD trial follow-up was censored at 5 years for the present analysis.

In the TECOS cohort, incident hospitalization for HF was adjudicated as a prespecified secondary end point of the trial. Incident hospitalization for HF was defined as an inpatient admission or emergency department visit >12 hours with clinical manifestations of HF and additional treatment with a diuretic, inotrope, vasodilator therapy, or mechanical or surgical intervention for hemodynamic support.[23] All events were prospectively collected and centrally adjudicated by an independent clinical committee of physicians masked to treatment assignment. The median follow-up was 3.0 (IQR, 2.3–3.7) years, and follow-up was censored at 4 years.

In the EHR cohort, the first episode of HF hospitalization was considered as an incident HF event. Time to the first HF event was calculated from the latest clinic visit date before December 31, 2014 (start of the outcome assessment period) to the first hospitalization with HF recorded as a physician-entered diagnosis on the hospital encounter form. Diagnoses were selected from the clinical terms available within the EHR (Intelligent Medical Objects, Chicago, IL) that were mapped to both Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) and *International Classification of Diseases, Ninth Revision*, and *International Classification of Diseases, Tenth Revision*, codes, as previously described.[24] HF was defined with SNOMED CT hierarchies for HF using the specific codes listed in Tables S1 and S2. All patient outcomes were censored at December 31, 2019, or at 5 years after the latest clinic visit date before December 31, 2014. The median follow-up was 5.7 (IQR, 4.0–6.0) years.

WATCH-DM Risk Score

The details and derivation of the WATCH-DM risk score have been previously published.[10] Briefly, the WATCH-DM risk score was developed to predict risk of incident HF and incorporates clinical, laboratory, and ECG parameters. Depending on the clinical use case, separate integer-, regression-, and machine learning (ML)–based models (henceforth referred to as WATCH-DM[i], WATCH-DM[r], and WATCH-DM[ml], respectively) were developed. Specifically, the integer-based model was designed to facilitate the ease of use in clinical settings without the need for a web-based platform or programming into the EHR system. The WATCH-DM score was developed using data from the ACCORD trial and validated in participants with T2D from the ALLHAT trial.[14,15] The risk score includes 10 variables, of which 3 are binary (QRS >120 milliseconds, history of MI, and history of CABG) and 7 are continuous (body mass index, age, systolic blood pressure, diastolic blood pressure, serum creatinine, high-density lipoprotein cholesterol, and FPG). For the present analyses, because ECG data were not available in Look AHEAD trial National Institute of Diabetes and Digestive and Kidney Disease data set and ECGs and FPG were not available in TECOS, the WATCH-DM risk score was rederived excluding QRS duration and replacing FPG for hemoglobin A1c using the same methods as the original score and described in the Supplemental Methods (Data S1). The rederived WATCH-DM risk integer score is shown in Figure S1. The performance of the rederived WATCH-DM risk scores is shown in Figure S2. To facilitate clinical use of the WATCH-DM(ml) model, prevent sharing of protected health data, and adhere to data use agreements, the model was implemented using an application programming interface, as detailed (Data S1).

TRS-HF_{DM} Risk Score

Details of the TRS-HF_{DM} risk score have been previously described.[11] TRS-HF_{DM} is an integer-based risk score to predict risk of HF hospitalization with the following variables and corresponding points: prior HF (2 points), history of atrial fibrillation (1 point), coronary artery disease (1 point), estimated glomerular filtration rate <60 mL/min per 1.73 m^2 (1 point), and urine albumin/creatinine ratio (UACR) >300 mg/g (2 points) or 30 to 300 mg/g (1 point). The risk score was developed using data from the SAVOR-TIMI 53 trial and validated in participants from the DECLARE-TIMI 58 and ACCORD trials.[17,25,26] All participants were assigned 0 points for prior HF because history of HF was an exclusion criterion for the present study.

Statistical Analysis

Baseline characteristics for each of the cohorts were compared across quintiles of WATCH-DM(i) and categories of TRS-HF_{DM} scores (0, 1, 2, and ≥3 point categories) and summarized as median (25th–75th percentiles) for continuous and number (percentage) for categorical variables. The unadjusted 5-year risk of incident HF was estimated in the EHR and Look AHEAD trial cohorts using Kaplan-Meier curves. Because of shorter follow-up in TECOS, the 4-year risk of incident HF was calculated. In addition, given UACR was missing in 63.4% of participants in TECOS,

analysis of TRS-HF_{DM} was restricted to only those with available UACR data (n=4408). The cumulative risk of incident HF was assessed across WATCH-DM(i) quintiles and TRS-HF_{DM} categories. Model performance was evaluated according to discrimination, assessed by the Harrell concordance index (C-index). Calibration was assessed by the Greenwood-Nam-D'Agostino method, with adequate calibration defined a priori as Greenwood-Nam-D'Agostino $P>0.05$.[27–29] Unadjusted Cox proportional hazard models were constructed to evaluate the association of score categories with risk of incident HF. Sensitivity analyses were also conducted to (1) compare predicted versus observed event rates across original WATCH-DM(i) score categories[10]; and (2) assess the performance of the WATCH-DM(ml) model in a less restrictive EHR cohort (ie, not requiring patients to have variables collected within 12 months of the baseline visit and not excluding patients because of missing variables).

Performance of WATCH-DM(i) was compared with TRS-HF_{DM} risk score in the TECOS and Look AHEAD trial cohorts using the same model performance metrics discussed previously. TRS-HF_{DM} performance was unable to be assessed in the EHR cohort because of lack of UACR data. Decision curve analysis, a measure between the number of true-positive cases identified without an increase in false-positive rate, was performed to compare the clinical net benefit between models.[30] Given the lack of a consensus patient risk threshold for HF treatment, harm was removed from the decision curve calculation and risk was assessed at the cohort-specific event rate.

Analyses were performed using either R version 3.6.3 (R Foundation, Vienna, Austria) for the Look AHEAD trial and EHR analyses or SAS version 14.3 (SAS Institute, Cary, NC) for the TECOS analysis, with a 2-sided $P<0.05$ indicating statistical significance.

RESULTS

Baseline characteristics of the participants stratified by cohort are shown in Table 1. Participants in the TECOS were older and more likely to be men (Table 1). Among the 3 study cohorts, participants in the TECOS also had a higher average blood pressure, longer duration of diabetes diagnosis, higher serum creatinine, and higher prevalence of prior MI and CABG. Conversely, participants of the Look AHEAD trial were younger, had the lowest percentage of men, had the lowest baseline blood pressure and serum creatinine, and had the lowest prevalence of prior MI and CABG. Patients in the EHR cohort had a median age of 60 years, half were men (49.8%), and the group had the lowest percentage of self-reported White race (54.6%) and highest median hemoglobin A1c (7.3%) (Table 1). HF incidence rate was 7.5, 3.9, and 4.1 per 1000 person-years in the TECOS, Look AHEAD trial, and electronic health record cohorts, respectively.

Table 1. Baseline Characteristics of Participants in the Look AHEAD Trial, TECOS, and EHR Cohorts

Characteristic	Look AHEAD trial	TECOS	EHR
Total No.	4867	12 028	7475
Age, y	59 (55–63)	65 (59–71)	60 (50–68)
Men	2022 (41.5)	8668 (72.1)	3724 (49.8)
White race	3228 (66.3)	7761 (64.5)	4083 (54.6)
Black race	795 (16.3)	399 (3.3)	1650 (22.1)
Others or unknown	844 (17.3)	3868 (32.2)	1742 (23.3)
Body mass index, kg/m^2	34.9 (31.5–39.4)	29.1 (26.0–32.8)	30.8 (26.6–36.3)
Systolic BP, mm Hg	129 (117–141)	133 (124–145)	131 (120–145)
Diastolic BP, mm Hg	70 (64–77)	78 (70–83)	76 (68–83)
Diabetes duration, y	5 (2–10)	10 (5–16)	…
Serum creatinine, mg/dL	0.8 (0.7–0.9)	1.0 (0.8–1.1)	0.9 (0.7–1.2)
HDL-c, mg/dL	42 (35–50)	42 (35–50)	45 (37–56)
HbA1c, %	7.1 (6.4–7.9)	7.2 (6.8–7.7)	7.3 (6.5–7.9)
Prior MI	287 (5.9)	4667 (38.8)	992 (13.3)
Prior CABG	119 (2.4)	2946 (24.5)	592 (7.9)
Insulin use	896 (18.4)	2639 (21.9)	…
WATCH-DM(i) score	13 (11–15)	14 (12–17)	10 (8–13)
TRS-HF_{DM} score	0 (0–1)	1 (1–2)	…

Values are displayed as median (25th–75th percentiles) for continuous and number (percentage) for categorical variables. BP indicates blood pressure; CABG, coronary artery bypass grafting; EHR, electronic health record; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; Look AHEAD, Look Action for Health in Diabetes; MI, myocardial infarction; TECOS, Trial Evaluating Cardiovascular Outcomes With Sitagliptin; TRS-HF_{DM}, Thrombolysis in Myocardial Infarction (TIMI) risk score for heart failure in diabetes; and WATCH-DM(i), integer-based weight (body mass index), age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control (fasting plasma glucose), ECG QRS duration, myocardial infarction, and coronary artery bypass grafting.

Table 2. Discrimination and Calibration Metrics of the WATCH-DM(i), WATCH-DM(r), WATCH-DM(ml), and TRS-HF$_{DM}$ Scores for Predicting Risk of Incident HF in Each Cohort Analyzed

Variable	Look AHEAD trial		TECOS*		EHR	
	C-index (95% CI)	GND *P* value	C-index (95% CI)	GND *P* value	C-index (95% CI)	GND *P* value
WATCH-DM(i)	0.70 (0.64–0.76)	0.39	0.65 (0.61–0.68)	<0.001	0.73 (0.69–0.77)	0.96
WATCH-DM(r)	0.73 (0.67–0.78)	0.16	0.67 (0.63–0.70)	<0.001	0.73 (0.69–0.78)	0.27
WATCH-DM(ml)	0.76 (0.70–0.82)	0.61	0.63 (0.59–0.67)	<0.001	0.77 (0.73–0.80)	0.42
TRS-HF$_{DM}$	0.70 (0.65–0.75)	0.84	0.66 (0.60–0.72)	<0.001	Could not be assessed	

The 5-year risk of HF was assessed in the Look AHEAD trial and EHR cohorts and 4-year risk in the TECOS cohort. C-index indicates concordance index; EHR, electronic health record; GND, Greenwood-Nam-D'Agostino; HF, heart failure; Look AHEAD, Look Action for Health in Diabetes; TECOS, Trial Evaluating Cardiovascular Outcomes With Sitagliptin; TRS-HF$_{DM}$, Thrombolysis in Myocardial Infarction (TIMI) risk score for heart failure in diabetes; WATCH-DM, weight (body mass index), age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control (fasting plasma glucose), ECG QRS duration, myocardial infarction, and coronary artery bypass grafting; WATCH-DM(i), integer-based WATCH-DM; WATCH-DM(ml), machine learning–based WATCH-DM; and WATCH-DM(r), regression-based WATCH-DM.

*Because of limited availability of urine albumin/creatinine ratio data, risk score performance in TECOS was assessed in 12 028 participants for WATCH-DM and 4408 participants in TECOS.

Performance of WATCH-DM and TRS-HF$_{DM}$ in the Look AHEAD Trial Cohort

Among 4867 participants from the Look AHEAD trial, 91 developed HF within 5 years from enrollment, with a 5-year Kaplan-Meier risk estimate of 1.83%. The median WATCH-DM(i) score was 13 (IQR, 11–15), with an observed range of 3 to 26. Baseline characteristics of participants stratified by WATCH-DM(i) quintiles are shown in Table S3. The WATCH-DM(i) score demonstrated good discrimination, with a C-index of 0.70 (95% CI, 0.64–0.76) and adequate calibration (*P*=0.39) for predicting HF risk in the Look AHEAD trial (Table 2

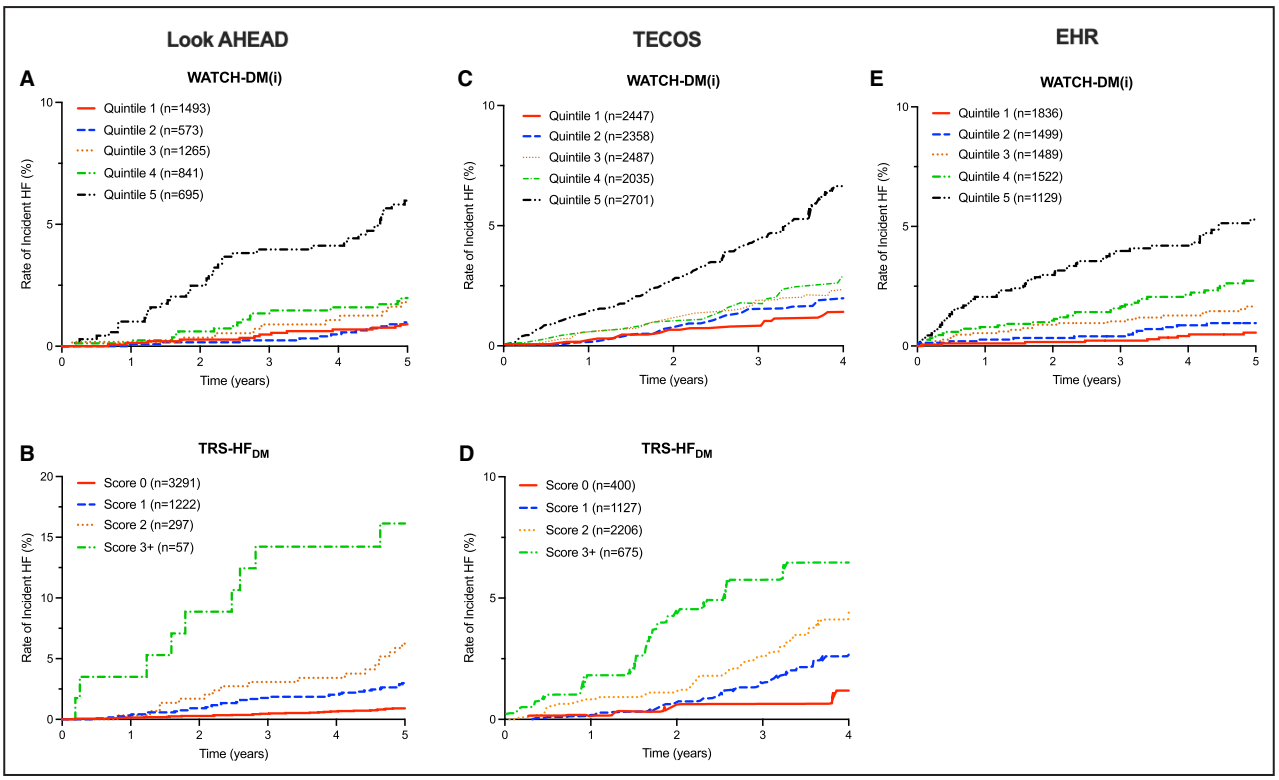


Figure 1. Cumulative incidence of heart failure (HF) in the Look AHEAD (Look Action for Health in Diabetes) trial (A and B), TECOS (Trial Evaluating Cardiovascular Outcomes With Sitagliptin) (C and D), and electronic health record (EHR) (E) validation cohorts across WATCH-DM(i) (integer-based weight [body mass index], age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control [fasting plasma glucose], ECG QRS duration, myocardial infarction, and coronary artery bypass grafting) and TRS-HFDM (Thrombolysis in Myocardial Infarction (TIMI) risk score for heart failure in diabetes) risk scores.

The 5-year risk was assessed in the Look AHEAD trial and EHR cohorts and 4-year risk in the TECOS cohort.

and Figure S3). Similar results were observed across original WATCH-DM(i) score categories (Figure S4). The 5-year incidence of HF increased across quintiles of WATCH-DM(i), ranging from 0.90% in quintile 1 to 5.97% in quintile 5 (Figure 1A). Event rates across quintiles are shown in Table S4. The risk of incident HF was almost 7-fold higher in quintile 5 compared with quintile 1 (hazard ratio [HR], 6.79; 95% CI, 3.32–15.27) (Table S5). The WATCH-DM(ml) and WATCH-DM(r) scores demonstrated superior discrimination, with C-indexes of 0.76 (95% CI, 0.70–0.82) and 0.73 (95% CI, 0.67–0.78), respectively (Table 2), and adequate calibration (P=0.61 and P=0.16, respectively) (Figure S3).

The median TRS-HF_{DM} score was 0 (IQR, 0–1), with an observed range of 0 to 1. Baseline characteristics of participants stratified by TRS-HF_{DM} are shown in Table S6. The TRS-HF_{DM} score demonstrated similar performance to the WATCH-DM(i) score, with a C-index of 0.70 (95% CI, 0.65–0.75) and adequate calibration (P=0.84) (Table 2 and Figure S5A). The incidence of HF increased from 0.91% for score of 0 to 16.10% for scores of ≥3 (Figure 1B and Table S4). The risk of incident HF was significantly higher for participants with scores of ≥3 compared with score of 0 (HR, 20.04; 95% CI, 9.49–42.34) (Table S5).

In decision curve analysis, WATCH-DM(i) identified 2 additional HF events per 1000 participants compared with the TRS-HF_{DM} risk score (Figure 2A).

Performance of WATCH-DM and TRS-HF_{DM} in the TECOS cohort

In 12 028 participants from the TECOS, 266 developed HF during follow-up, with a 4-year Kaplan-Meier risk of 3.1%. The median WATCH-DM(i) score was 14 (IQR, 12–17), with an observed range of 3 to 30. Baseline characteristics of participants across quintiles of WATCH-DM(i) are shown in Table S7. The WATCH-DM(i), WATCH-DM(r), and WATCH-DM(ml) models demonstrated modest discrimination for predicting risk of HF in the TECOS, with C-indexes of 0.65 (95% CI, 0.61–0.68), 0.67 (95% CI, 0.63–0.70), and 0.63 (95% CI, 0.59–0.67), respectively (Table 2). Evidence of miscalibration was observed (P<0.001) in all models particularly underpredicting observed risk in the highest deciles and categories (Figures S3 through S4). The WATCH-DM(i) score was able to stratify the incidence of HF at year 4 from 1.85% in quintile 1 to 6.66% in quintile 5 (Figure 1C and Table S4). Similarly, the risk of incident HF was nearly 4-fold higher for quintile 5 when compared with quintile 1 (HR, 3.78; 95% CI, 2.54–5.61) (Table S5).

Among 4408 participants with available UACR data, the median TRS-HF_{DM} score was 1 (IQR, 1–2), with an observed range of 0 to 5. Baseline characteristics across score categories are shown in Table S8. The TRS-HF_{DM} score demonstrated modest discrimination, with a C-index of 0.66 (95% CI, 0.60–0.72) for predicting HF risk in the TECOS cohort (Table 2). Evidence of miscalibration was observed with

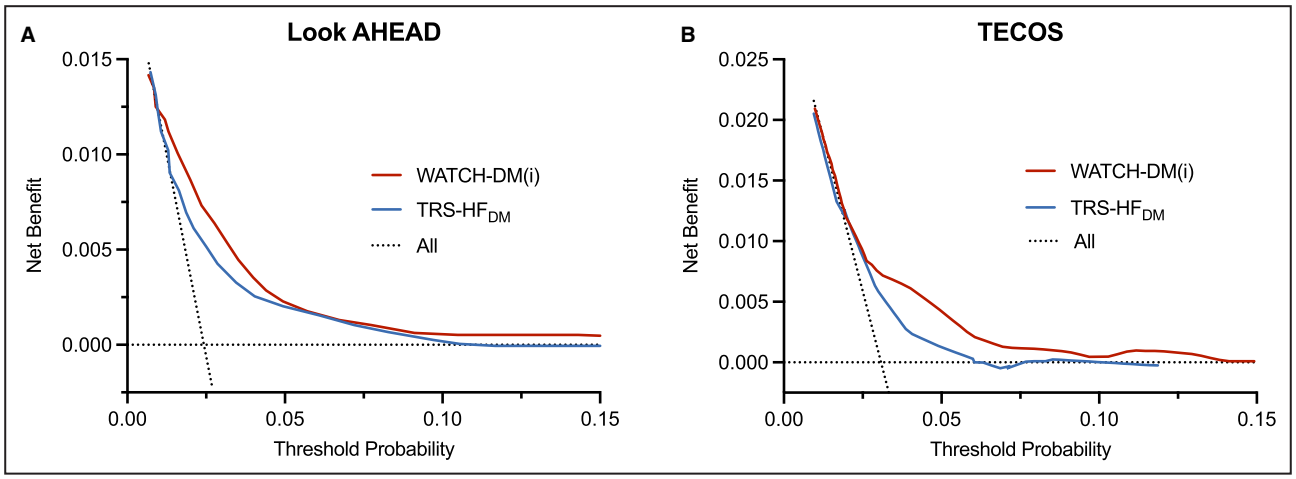


Figure 2. Decision curve analysis of the WATCH-DM(i) (integer-based weight [body mass index], age, hypertension, creatinine, high-density lipoprotein cholesterol, diabetes control [fasting plasma glucose], ECG QRS duration, myocardial infarction, and coronary artery bypass grafting) and TRS-HFDM (Thrombolysis in Myocardial Infarction (TIMI) risk score for heart failure in diabetes) risk scores in the Look AHEAD (Look Action for Health in Diabetes) trial (A) and TECOS (Trial Evaluating Cardiovascular Outcomes With Sitagliptin) (B) validation cohorts.
At a 1.9% risk threshold (the overall heart failure event rate) in the Look AHEAD trial cohort, the WATCH-DM(i) risk score identified 2 additional heart failure events per 1000 individuals compared with the TRS-HF_{DM} risk score. Similarly, at a 3% risk threshold in the TECOS cohort, the WATCH-DM(i) risk score identified 2 additional heart failure events per 1000 individuals compared with the TRS-HF_{DM} risk score.

consistent underestimated HF risk and a Greenwood-Nam-D'Agostino $P<0.001$ (Table 2 and Figure S5B). The incidence of HF increased from 1.16% for a score of 0 to 6.42% for scores of ≥3 (Figure 1D and Table S4). Compared with participants with a score of 0, the risk of incident HF was >6-fold higher for scores of ≥3 (HR, 6.71; 95% CI, 2.71–16.60) (Table S5). In decision curve analysis, WATCH-DM(i) identified 2 additional HF events per 1000 individuals compared with the TRS-HF_{DM} risk score (Figure 2B).

External Validation of WATCH-DM in the EHR Cohort

Among 7475 patients from the UT Southwestern Medical Center EHR cohort, 133 developed incident HF within 5 years from the baseline clinic visit, with a Kaplan-Meier risk of 1.88%. The median WATCH-DM(i) score was 10 (IQR, 8–13), with an observed range of 1 to 28. Baseline characteristics of participants stratified by WATCH-DM(i) quintiles are shown in Table S9. The WATCH-DM(i) score demonstrated good discrimination, with a C-index of 0.73 (95% CI, 0.69–0.77) and adequate calibration ($P=0.96$) (Table 2 and Figures S3 and S4). The incidence of HF increased across quintiles of WATCH-DM(i), ranging from 0.56% in quintile 1 to 5.29% in quintile 5 (Figure 1E and Table S4). The risk of incident HF was nearly 9-fold higher in quintile 5 compared with quintile 1 (HR, 8.98; 95% CI, 4.46–18.06) (Table S5). The WATCH-DM(r) and WATCH-DM(ml) scores also demonstrated good discrimination, with C-indexes of 0.73 (95% CI, 0.69–0.78) and 0.77 (95% CI, 0.73–0.80) (Table 2), and adequate calibration ($P=0.27$ and $P=0.42$, respectively) (Figure S3). In sensitivity analysis, we liberalized the cohort to include patients with laboratory values collected at different times (ie, >12 months of the baseline visit) or patients with >20% variable missingness. The analysis cohort included 13 516 patients (75.4% of candidate cohort). Among 13 516 patients, 8947 (66.2%) were missing variables required for WATCH-DM (average participant variable missingness, 10% [IQR, 0%–30%]). Even in the presence of considerable missingness, the WATCH-DM(ml) score demonstrated modest discrimination, with C-index of 0.70 (95% CI, 0.67–0.72), and adequate calibration ($P=0.11$).

DISCUSSION

In the present study, we evaluated the performance of the novel HF risk scores for T2D, the WATCH-DM score and the TRS-HF_{DM} score, in large external cohorts of individuals with varying cardiovascular risk profiles. Among adults with T2D and intermediate cardiovascular disease risk from the Look AHEAD trial cohort, both risk scores predicted 5-year risk of incident HF with adequate discrimination and calibration. Conversely, among the high-risk cohort of adults with T2D in the TECOS, neither WATCH-DM nor TRS-HF_{DM} was well calibrated and tended to underestimate HF risk. The WATCH-DM risk score detected an additional 2 HF events per 1000 individuals compared with the TRS-HF_{DM} risk score in both the Look AHEAD trial and TECOS cohorts. Finally, we demonstrated that the WATCH-DM risk score can be calculated from data acquired from routine clinical care in the EHR.

Over the past 2 decades, among adults with T2D, there has been substantial progress in reducing hospitalizations for ischemic heart disease but less so for HF.[3] Primary cardiovascular disease prevention guidelines endorse use of the pooled cohort equations to estimate 10-year risk of atherosclerotic cardiovascular disease to inform allocation of preventive therapies to reduce risk, but there is no specific instrument recommended for HF risk assessment to similarly inform clinical decision making.[31] Multiple HF risk prediction tools are available, but none has been widely accepted because, in part, of the need for robust validation.[32] The pooled cohort equations to prevent HF were developed in pooled epidemiologic cohorts to predict incident HF and were validated in a low-risk cohort free of cardiovascular disease.[13,29] However, these prior HF risk prediction equations were not specific for T2D and excluded patients with prevalent coronary heart disease, a common comorbidity and HF risk factor among patients with T2D, thereby limiting their generalizability. The integer-based TRS-HF_{DM} risk score was developed using clinical trial data to predict HF hospitalization in T2D and was externally validated in a separate trial cohort.[11,17] However, the TRS-HF_{DM} risk score requires UACR to estimate risk, but this is not routinely assessed in clinical practice, limiting its use.

In the present study, we evaluated the performance of the WATCH-DM(i) and TRS-HF_{DM} risk scores in cohorts with intermediate (Look AHEAD trial) and high baseline cardiovascular risk (TECOS). Both the WATCH-DM and TRS-HF_{DM} scores identified a wide gradient of risk across categories with good discrimination across cohorts with variable baseline cardiovascular disease risk. Both scores also had adequate calibration in the intermediate- but not high-risk cohort. In the high-risk cohort, the WATCH-DM risk score underestimated HF risk, especially in the highest deciles of risk. For WATCH-DM, miscalibration was likely related to differences in baseline cardiovascular disease risk among the derivation (ACCORD) and validation study (TECOS) populations. Notably, the TECOS cohort included primarily men (72%) with high risk for developing atherosclerotic cardiovascular disease. Recalibration of the contribution of history of MI and history of CABG in a higher-risk cohort or including other complications

of T2D as variables, such as cerebrovascular disease or peripheral vascular disease, may yield improved results. The TRS-HF_{DM} risk score underestimated HF risk in TECOS compared with lower-risk cohort, such as the Look AHEAD trial and ACCROD trial cohorts.[17] Overall, both risk scores may have better performance in lower- than in higher-risk populations. Even in the SAVOR-TIMI cohort and the DECLARE-TIMI cohort, where the TIMI-HF_{DM} was developed, was initially validated, and demonstrated good performance, the proportion of participants with prior MI, peripheral vascular disease, and cerebrovascular disease was lower than that observed in the TECOS validation cohort. These differences in ischemic vascular disease and associated risk of HF may explain the variability in performance of TRS-HF_{DM} across cohorts.[11] It is also noteworthy that TRS-HF_{DM} risk score was developed to predict HF hospitalization among patients with diabetes with or without history of HF, with prevalent HF contributing 2 points (of maximum 8 points) toward the risk score.[11] In the present study, we used TRS-HF_{DM} to predict incident HF with 0 points assigned to the history of HF criteria, which may have further limited its performance in TECOS compared with other high-risk cohorts, such as DECLARE-TIMI cohort.[11]

There are several strengths to the WATCH-DM risk score with respect to its implementation for management of patients with T2D in the contemporary clinical setting.

We demonstrated easy applicability of the risk score in a health system's EHR for efficient risk estimation in clinical practice. WATCH-DM could be calculated for most individuals without significant exclusions required for data missingness. Comparatively, because of lack of available UACR data, the TRS-HF_{DM} score could not be calculated in the EHR cohort and only in <40% of participants in TECOS. Clinical decision support tools in prevalent HF are associated with improved medication adherence and appropriate referrals for advanced therapies, but less is known about their impact on HF prevention.[33,34] In addition to the feasibility from an implementation standpoint, WATCH-DM demonstrated adequate discrimination and calibration for predicting HF risk in a real-world contemporary clinical cohort of patients with T2D. These observations highlight the potential for wide generalizability and applicability of the WATCH-DM risk score. Participants included in the EHR cohort are not subject to strict enrollment criteria that often challenge the generalizability of findings from clinical trials.[35] In addition, the clinical data used to estimate 5-year incident HF risk were obtained as part of routine clinical care rather than standardized protocols used in research studies.

ML models also afford several advantages compared with traditional risk modeling techniques. Notably, ML models can be updated to tailor to the available data, making them more usable across different populations. Similar to other maximum likelihood-based modeling approaches, ML models can continue to function even in the presence of considerable missing data, as would be common in real-world or EHR data registries. In the present study, we observed that the WATCH-DM(ml) model performed well even in an EHR cohort with 65% of patients missing the required variables. In addition, we validated a proof-of-concept application programming interface to share the WATCH-DM(ml) model while preserving data use agreements. Such a model allows for WATCH-DM(ml) to be directly accessed by researchers or applications without directly sharing patient-level data.

The present study findings have important clinical implications. Clinical practice recommendations suggest select individuals with T2D should be considered for weight loss therapies, including intensive lifestyle interventions and metabolic surgery, as well as specific medications, such as SGLT2i, based on comorbidity burden and risk for HF.[36–38] Intentional weight loss, particularly targeting central adiposity, and prescription of SGLT2i are associated with lower risk of HF, but, currently, there are no validated risk scores recommended for HF risk stratification.[8,39,40] Both WATCH-DM and TRS-HF_{DM} are HF risk prediction tools that incorporate routinely assessed clinical data and are now validated in multiple cohorts. In the present analyses, we demonstrated that novel risk scores could help target preventive HF therapies, such as intentional weight loss interventions and SGLT2i, to individuals who have the highest risk for developing HF and are therefore most likely to experience the greatest absolute risk reductions for incident HF. Clinical decision support tools incorporating risk assessment may help target HF preventive therapies to individuals at the greatest risk for incident HF in a cost-effective manner. In our study, both the WATCH-DM and TRS-HF_{DM} risk scores had relatively worse performance in the high-risk TECOS cohort. However, high-risk patients, such as those in the TECOS cohort, are less in need for risk stratification and would benefit from preventive therapies because of high baseline event rate. The application of risk stratification models, such as WATCH-DM and TRS-HF_{DM}, is more relevant for low- and intermediate-risk populations, where these scores demonstrated adequate to good performance. Future studies are needed to determine if implementing these risk score in the EMR may improve uptake of evidence-based therapies for HF prevention in patients with T2D.

This study has several notable strengths, including comparison of 2 novel T2D-specific risk scores, validation in 2 large clinical trial cohorts with study populations who have varied baseline cardiovascular disease risk, additional validation in a real-world EHR data set, inclusion of a diverse study population, and rigorous adjudication of HF events in the clinical

trial cohorts, according to standardized protocols. However, the study findings should be interpreted in the context of several limitations. First, the WATCH-DM risk score assessed in the present study was modified from its original derivation.[10] ECGs and fasting blood samples are not routinely collected in clinical practice. The WATCH-DM risk score was rederived excluding ECG data and substituting hemoglobin A1c for FPG using similar methods as the original analysis to ensure generalizability of the risk prediction tool in clinical practice. The rederived WATCH-DM risk score was validated and had similar model performance for predicting HF risk as the original score (C-index range, 0.72–0.76; Figure S2)[10]; however, with fewer variables, the ML-based model was more prone to overfitting. Second, data from the EHR were collected according to routine clinical practice. WATCH-DM was assessed in the real-world EHR data set without strict data requirements, suggesting that this risk score may have generalizable use and implementation may be feasible. Also, UACR data were not captured consistently in the EHR data set to allow for evaluation of the TRS-HF_{DM} risk score. Similarly, <40% of participants in TECOS had available UACR data. As such, evaluation of TRS-HF_{DM} was limited to only those with available data. Third, because of the required ECG variables, we were unable to assess the performance of the pooled cohort equations to prevent HF risk score.13 Finally, both risk scores do not incorporate blood-based biomarkers, such as high-sensitivity cardiac troponin and natriuretic peptide levels, that are well-established predictors of HF risk.[41–43] Similarly, data on cardiometabolic measures, such as fat mass, cardiorespiratory fitness, and visceral adiposity, which are also well-established risk factors of HF, were not included in either risk score.[40,44,45] These factors are not commonly assessed in routine clinical practice and thus are not readily available in all patients with T2D for risk assessment. Future studies are needed to determine whether incorporation of these markers of risk can improve the predictive performance of the WATCH-DM and TRS-HF_{DM} risk scores.

In conclusion, the WATCH-DM and TRS-HF_{DM} risk scores were able to stratify HF risk among adults with T2D and differing degrees of baseline cardiovascular disease risk in 2 large, multicenter, clinical trials. The WATCH-DM risk score was also validated in a contemporary EHR cohort from a large health system. Future studies are needed to further evaluate whether use of the WATCH-DM or TRS-HF_{DM} risk scores can improve use of effective preventive interventions, such as SGLT2i, to lower the risk of HF.

ARTICLE INFORMATION

Received October 15, 2021; accepted March 28, 2022.

Affiliations

Department of Cardiology, Texas Heart Institute, Houston, TX (M.W.S.); Department of Cardiology, Houston Methodist DeBakey Heart and Vascular Center, Houston, TX (K.V.P.); Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC (A.S.H., Y.L., J.B.G., E.D.P.); Brigham and Women's Hospital Heart and Vascular Center, Department of Medicine, Harvard Medical School, Boston, MA (M.V.); Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX (S.-H.W., A.A.K., E.D.P., V.K., D.L.W., D.K.M., A.P.); Diabetes Trials Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, UK (R.R.H.); and Parkland Health and Hospital System, Dallas, TX (E.D.P., D.K.M.).

Acknowledgments

The Look AHEAD (Look Action for Health in Diabetes) trial was conducted by the Look AHEAD Research Group and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK); the National Heart, Lung, and Blood Institute; the National Institute of Nursing Research; the National Institute of Minority Health and Health Disparities; the Office of Research on Women's Health; and the Centers for Disease Control and Prevention. The data (and samples) from Look AHEAD trial were supplied by the NIDDK Central Repository. This article was not prepared under the auspices of the Look AHEAD trial and does not represent analyses or conclusions of the Look AHEAD Research Group, the NIDDK Central Repository, or the National Institutes of Health.

Sources of Funding

The TECOS (Trial Evaluating Cardiovascular Outcomes With Sitagliptin) was funded by Merck Sharp & Dohme Corp, a subsidiary of Merck & Co, Inc, Kenilworth, NJ. Dr Vaduganathan is supported by the KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst (National Institutes of Health/National Center for Advancing Translational Sciences Award UL 1TR002541) and serves on advisory boards or has received research grant support from American Regent, Amgen, AstraZeneca, Baxter Healthcare, Bayer AG, Boehringer Ingelheim, Cytokinetics, and Relypsa. Dr Holman reports research support from AstraZeneca, Bayer, and Merck Sharp & Dohme, and personal fees from Anji Pharmaceuticals, AstraZeneca, Novartis, and Novo Nordisk. Dr. Kolkailah was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number T32HL125247. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health. Dr McGuire has had leadership roles in clinical trials for AstraZeneca, Boehringer Ingelheim, Eisai, Esperion, GlaxoSmithKline, Janssen, Lexicon, Merck & Co, Inc, Novo Nordisk, CSL Behring, and Sanofi USA; and has received consultancy fees from AstraZeneca, Boehringer Ingelheim, Lilly USA, Merck & Co, Inc, Pfizer, Novo Nordisk, Metavant, Afimmune, and Sanofi. Dr Pandey received grant funding outside the present study from Applied Therapeutics; has received honoraria outside of the present study as an advisor/consultant for Tricog Health Inc, Lilly, USA, Rivus, and Roche Diagnostics; and has received nonfinancial support from Pfizer and Merck. Dr Pandey is supported by the Texas Health Resources Clinical Scholarship, Gilead Sciences Research Scholar Program, the National Institute of Aging GEMSSTAR Grant (1R03AG067960-01), and grant support from Applied Therapeutics. Dr Peterson receives consulting from: advisory committees for Novo Nordisk, Novartis, Janssen, Pfizer, Bayer; and receives research support from: Janssen, Amgen, Esperion, BMS

Disclosures

All other authors report no disclosures or sources of funding.

Supplemental Material

Data S1
Tables S1–S9
Figures S1–S5
References 46–49

REFERENCES

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. Heart disease and stroke statistics-2020 update: a report from the American Heart Association. *Circulation*. 2020;141:e139–e596.

2. Cavender MA, Steg PG, Smith SC Jr, Eagle K, Ohman EM, Goto S, Kuder J, Im K, Wilson PW, Bhatt DL, et al. Impact of diabetes mellitus on hospitalization for heart failure, cardiovascular events, and death: outcomes at 4 years from the Reduction of Atherothrombosis for Continued Health (REACH) Registry. *Circulation*. 2015;132:923–931. doi: 10.1161/CIRCULATIONAHA.114.014796
3. Honigberg MC, Patel RB, Pandey A, Fonarow GC, Butler J, McGuire DK, Vaduganathan M. Trends in hospitalizations for heart failure and ischemic heart disease among US adults with diabetes. *JAMA Cardiol*. 2021;6:354–357. doi: 10.1001/jamacardio.2020.5921
4. Khera R, Kondamudi N, Zhong L, Vaduganathan M, Parker J, Das SR, Grodin JL, Halm EA, Berry JD, Pandey A. Temporal trends in heart failure incidence among medicare beneficiaries across risk factor strata, 2011 to 2016. *JAMA Network Open*. 2020;3(10):e2022190. doi: 10.1001/jamanetworkopen.2020.22190
5. Dauriz M, Targher G, Laroche C, Temporelli PL, Ferrari R, Anker S, Coats A, Filippatos G, Crespo-Leiro M, Mebazaa A, et al. Association between diabetes and 1-year adverse clinical outcomes in a multinational cohort of ambulatory patients with chronic heart failure: results from the ESC-HFA heart failure long-term registry. *Diabetes Care*. 2017;40:671–678. doi: 10.2337/dc16-2016
6. Rawshani A, Rawshani A, Franzén S, Sattar N, Eliasson B, Svensson A-M, Zethelius B, Miftaraj M, McGuire DK, Rosengren A, et al. Risk factors, mortality, and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2018;379:633–644. doi: 10.1056/NEJMoa1800256
7. Wright AK, Suarez-Ortegon MF, Read SH, Kontopantelis E, Buchan I, Emsley R, Sattar N, Ashcroft DM, Wild SH, Rutter MK. Risk factor control and cardiovascular event risk in people with type 2 diabetes in primary and secondary prevention settings. *Circulation*. 2020;142:1925–1936. doi: 10.1161/CIRCULATIONAHA.120.046783
8. McGuire DK, Shih WJ, Cosentino F, Charbonnel B, Cherney DZI, Dagogo-Jack S, Pratley R, Greenberg M, Wang S, Huyck S, et al. Association of SGLT2 inhibitors with cardiovascular and kidney outcomes in patients with type 2 diabetes: a meta-analysis. *JAMA Cardiol*. 2020;6:148. doi: 10.1001/jamacardio.2020.4511
9. Vaduganathan M, Sathiyakumar V, Singh A, McCarthy CP, Qamar A, Januzzi JL Jr, Scirica BM, Butler J, Cannon CP, Bhatt DL. Prescriber patterns of SGLT2i after expansions of U.S. food and drug administration labeling. *J Am Coll Cardiol*. 2018;72:3370–3372.
10. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, Basit M, Kannan V, Grodin JL, Everett B, et al. Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score. *Diabetes Care*. 2019;42:2298–2306. doi: 10.2337/dc19-0587
11. Berg DD, Wiviott SD, Scirica BM, Gurmu Y, Mosenzon O, Murphy SA, Bhatt DL, Leiter LA, McGuire DK, Wilding JPH, et al. Heart failure risk stratification and efficacy of sodium-glucose cotransporter-2 inhibitors in patients with type 2 diabetes mellitus. *Circulation*. 2019;140:1569–1577. doi: 10.1161/CIRCULATIONAHA.119.042685
12. Berg DD, Wiviott SD, Scirica BM, Zelniker TA, Goodrich EL, Jarolim P, Mosenzon O, Cahn A, Bhatt DL, Leiter LA, A biomarker-based score for risk of hospitalization for heart failure in patients with diabetes. *Diabetes Care*. 2021;44(11):2573–2581. doi: 10.2337/dc21-1170
13. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, Mentz RJ, O'Brien E, Correa A, Suthahar N, et al. 10-year risk equations for incident heart failure in the general population. *J Am Coll Cardiol*. 2019;73:2388–2397.
14. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358:2545–2559.
15. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA*. 2002;288:2981–2997. doi: 10.1001/jama.288.23.2981
16. Segar MW, Khan MS, Patel KV, Vaduganathan M, Kannan V, Willett D, Peterson E, Tang WHW, Butler J, Everett BM, et al. Incorporation of natriuretic peptides with clinical risk scores to predict heart failure among individuals with dysglycaemia. *Eur J Heart Fail*. 2022;24:169–180. doi: 10.1002/ejhf.2375
17. Elharram M, Ferreira JP, Huynh T, Ni J, Giannetti N, Verma S, Zannad F, Sharma A. Prediction of heart failure outcomes in patients with type 2 diabetes mellitus: validation of the Thrombolysis in Myocardial Infarction Risk Score for Heart Failure in Diabetes (TRS-HFDM) in patients in the ACCORD trial. *Diabetes Obes Metab*. 2021;23:782–790.
18. Look Ahead Research Group, Wing RR, Bolin P, Brancati FL, Bray GA, Clark JM, Coday M, Crow RS, Curtis JM, Egan CM, et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013;369:145–154.
19. Green JB, Bethel MA, Paul SK, Ring A, Kaufman KD, Shapiro DR, Califf RM, Holman RR. Rationale, design, and organization of a randomized, controlled Trial Evaluating Cardiovascular Outcomes with Sitagliptin (TECOS) in patients with type 2 diabetes and established cardiovascular disease. *Am Heart J*. 2013;166:983–989.e7. doi: 10.1016/j.ahj.2013.09.003
20. Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, Josse R, Kaufman KD, Koglin J, Korn S, et al. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2015;373:232–242. doi: 10.1056/NEJMoa1501352
21. Ryan DH, Espeland MA, Foster GD, Haffner SM, Hubbard VS, Johnson KC, Kahn SE, Knowler WC, Yanovski SZ, Look ARG. Look AHEAD (Action for Health in Diabetes): design and methods for a clinical trial of weight loss for the prevention of cardiovascular disease in type 2 diabetes. *Control Clin Trials*. 2003;24:610–628. doi: 10.1016/S0197-2456(03)00064-3
22. Stekhoven DJ, Buhlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28:112–118. doi: 10.1093/bioinformatics/btr597
23. McGuire DK, Van de Werf F, Armstrong PW, Standl E, Koglin J, Green JB, Bethel MA, Cornel JH, Lopes RD, Halvorsen S, et al. Association between sitagliptin use and heart failure hospitalization and related outcomes in type 2 diabetes mellitus: secondary analysis of a randomized clinical trial. *JAMA Cardiol*. 2016;1:126–135. doi: 10.1001/jamacardio.2016.0103
24. Willett D, Kannan V, Chu L, Buchanan J, Velasco F, Clark J, Fish J, Ortuzar A, Youngblood J, Bhat D, et al. SNOMED CT concept hierarchies for sharing definitions of clinical conditions using electronic health record data. *Appl Clin Inform*. 2018;9:667–682. doi: 10.1055/s-0038-1668090
25. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, Ohman P, Frederich R, Wiviott SD, Hoffman EB, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med*. 2013;369:1317–1326. doi: 10.1056/NEJMoa1307684
26. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, Zelniker TA, Kuder JF, Murphy SA, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2019;380:347–357. doi: 10.1056/NEJMoa1812389
27. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med*. 2015;34:1659–1680. doi: 10.1002/sim.6428
28. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. *Circulation*. 2010;121:1768–1777. doi: 10.1161/CIRCULATIONAHA.109.849166
29. Bavishi A, Bruce M, Ning H, Freaney PM, Glynn P, Ahmad FS, Yancy CW, Shah SJ, Allen NB, Vupputuri SX, et al. Predictive accuracy of heart failure-specific risk equations in an electronic health record-based cohort. *Circ Heart Fail*. 2020;13:e007462. doi: 10.1161/CIRCHEARTFAILURE.120.007462
30. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26:565–574. doi: 10.1177/0272989X06295361
31. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140:e596–e646.
32. Echouffo-Tcheugui JB, Greene SJ, Papadimitriou L, Zannad F, Yancy CW, Gheorghiade M, Butler J. Population risk prediction models for incident heart failure: a systematic review. *Circ Heart Fail*. 2015;8:438–447. doi: 10.1161/CIRCHEARTFAILURE.114.001896
33. McKie PM, Kor DJ, Cook DA, Kessler ME, Carter RE, Wilson PM, Pencille LJ, Hickey BC, Chaudhry R. Computerized advisory decision support for cardiovascular diseases in primary care: a cluster randomized trial. *Am J Med*. 2020;133:750–756.e2. doi: 10.1016/j.amjmed.2019.10.039
34. Evans RS, Kfoury AG, Horne BD, Lloyd JF, Benuzillo J, Rasmusson KD, Roberts C, Lappe DL. Clinical decision support to efficiently

identify patients eligible for advanced heart failure therapies. *J Card Fail*. 2017;23:719–726. doi: 10.1016/j.cardfail.2017.08.449
35. Goldstein BA, Navar AM, Pencina MJ. Risk prediction with electronic health records: the importance of model validation and clinical context. *Jama Cardiol*. 2016;1:976–977. doi: 10.1001/jamacardio.2016.3826
36. American Diabetes A. 9. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S98–S110.
37. American Diabetes A. 8. Obesity management for the treatment of type 2 diabetes: standards of medical care in diabetes-2021. *Diabetes Care*. 2021;44:S100–S110.
38. Das SR, Everett BM, Birtcher KK, Brown JM, Januzzi JL, Kalyani RR, Kosiborod M, Magwire M, Morris PB, Neumiller JJ, et al. 2020 expert consensus decision pathway on novel therapies for cardiovascular risk reduction in patients with type 2 diabetes: a report of the American College of Cardiology solution set oversight committee. *J Am Coll Cardiol*. 2020;76:1117–1145. doi: 10.1016/j.jacc.2020.05.037
39. Pandey A, Patel KV, Bahnson JL, Gaussoin SA, Martin CK, Balasubramanyam A, Johnson KC, McGuire DK, Bertoni AG, Kitzman D, et al. Association of intensive lifestyle intervention, fitness, and body mass index with risk of heart failure in overweight or obese adults with type 2 diabetes mellitus: an analysis from the look AHEAD trial. *Circulation*. 2020;141:1295–1306. doi: 10.1161/CIRCULATIONAHA.119.044865
40. Patel KV, Bahnson JL, Gaussoin SA, Johnson KC, Pi-Sunyer X, White U, Olson KL, Bertoni AG, Kitzman DW, Berry JD, et al. Association of baseline and longitudinal changes in body composition measures with risk of heart failure and myocardial infarction in type 2 diabetes: findings from the look AHEAD trial. *Circulation*. 2020;142:2420–2430. doi: 10.1161/CIRCULATIONAHA.120.050941
41. Pandey A, Patel KV, Vongpatanasin W, Ayers C, Berry JD, Mentz RJ, Blaha MJ, McEvoy JW, Muntner P, Vaduganathan M, et al. Incorporation of biomarkers into risk assessment for allocation of antihypertensive medication according to the 2017 ACC/AHA high blood pressure guideline: a pooled cohort analysis. *Circulation*. 2019;140:2076–2088. doi: 10.1161/CIRCULATIONAHA.119.043337
42. Pandey A, Vaduganathan M, Patel KV, Ayers C, Ballantyne CM, Kosiborod MN, Carnethon M, DeFilippi C, McGuire DK, Khan SS, et al. Biomarker-based risk prediction of incident heart failure in pre-diabetes and diabetes. *JACC Heart Fail*. 2021;9(3):215–223.
43. Pandey A, Keshvani N, Ayers C, Correa A, Drazner MH, Lewis A, Rodriguez CJ, Hall ME, Fox ER, Mentz RJ, et al. Association of cardiac injury and malignant left ventricular hypertrophy with risk of heart failure in African Americans: the Jackson Heart Study. *JAMA Cardiol*. 2019;4:51–58. doi: 10.1001/jamacardio.2018.4300
44. Pandey A, Cornwell WK III, Willis B, Neeland IJ, Gao A, Leonard D, DeFina L, Berry JD. Body mass index and cardiorespiratory fitness in mid-life and risk of heart failure hospitalization in older age: findings from the Cooper Center Longitudinal Study. *JACC Heart Fail*. 2017;5:367–374.
45. Pandey A, Kondamudi N, Patel KV, Ayers C, Simek S, Hall ME, Musani SK, Blackshear C, Mentz RJ, Khan H, et al. Association between regional adipose tissue distribution and risk of heart failure among blacks. *Circ Heart Fail*. 2018;11:e005629. doi: 10.1161/CIRCHEARTFAILURE.118.005629

SUPPLEMENTAL MATERIAL

Data S1. Supplemental Methods

Look AHEAD cohort: study population

The details of the Look Action for Health in Diabetes (AHEAD) trial have previously been published.[18] In brief, in the Look AHEAD trial, the cardiovascular effects of an intensive lifestyle intervention focused on weight loss versus a diabetes support and education intervention were assessed among adults with type 2 diabetes mellitus (T2DM). Participants were enrolled between August 2001 and April 2004 and included individuals aged 45-76 years who were overweight or obese (body mass index [BMI] either ≥ 25 kg/m^2, or ≥ 27 kg/m^2 if taking insulin) and had T2DM with hemoglobin A1c (HbA1c) $\leq 11\%$, blood pressure (BP) $\leq 160/100$ mm Hg, and a plasma triglyceride level < 600 mg/dL. Exclusion criteria included individuals with type 1 diabetes, New York Heart Association class III or IV heart failure (HF), diseases affecting safety or limiting lifespan, or difficulties with adherence. The lifestyle intervention arm included weekly individualized and group counseling for the first 6 months with gradually decreasing frequency for the remainder of the trial. Participants in the intensive arm were also encouraged to maintain at least 7% weight loss, achieve ≥ 175 minutes/week of moderate-intensity physical activity, and were prescribed a restricted-calorie diet (1200-1800 kcal/day). The trial was stopped for futility in September 2012 after a median follow-up of 9.6 years.

Look AHEAD cohort: clinical covariates

Baseline data were obtained through in-person screening visits and prior to the beginning of intervention. Age, history of myocardial infarction (MI), and history of coronary artery bypass graft (CABG) were obtained through self-reported questionnaires. BMI was measured using a digital scale and stadiometer and calculated using weight in kilograms divided by height in meters squared. HbA1c, serum creatinine, and high density lipoprotein cholesterol (HDL-c) were measured by assays performed at the Look AHEAD Central Laboratory as previously described.[18] Systolic and diastolic BP were measured using the average of two seated measurements obtained using an automated device after 5 minutes of rest.

TECOS cohort: study population

The design and results of the Trial Evaluating Cardiovascular Outcomes with Sitagliptin (TECOS) have previously been published.[19, 20] In TECOS, the cardiovascular safety of sitagliptin, a dipeptidyl peptidase 4 inhibitor, was assessed among 14,671 participants with T2DM. Eligible participants were aged ≥ 50 years, had T2D with HbA1c between 6.5-8.0% on stable glucose-lowering medications, and had prevalent atherosclerotic cardiovascular disease (ASCVD; either major coronary artery disease, ischemic cerebrovascular disease, or peripheral artery disease). Participants with an estimated glomerular filtration rate < 30 mL/min/1.73 m^2; concurrent use of a dipeptidyl peptidase 4 inhibitor, rosiglitazone, or glucagon-like peptide-1 receptor agonist; or two or more episodes of hypoglycemia requiring medical assistance within the previous 12 months were excluded.

TECOS cohort: clinical covariates

Detailed demographic, social, clinical, and laboratory information were collected at the time of study enrollment. Age, history of MI, and history of CABG were self-reported. Blood pressure, height, and weight were obtained per local clinic protocols. BMI was calculated as previously described. Blood samples for HbA1c were collected and measured at designated local laboratories. Serum creatinine and HDL-c levels were obtained from the participant's usual-care provider.

Electronic health record cohort: clinical covariates

Birthdate, sex, and race were self-reported during patient registration. Age was defined as the time between the baseline visit and birthdate. Systolic and diastolic BP, height, and weight were recorded in flowsheet rows and entered by members of the clinical care team. History of MI and CABG was obtained from recordings in the medical history or problem list in the electronic health record (EHR). The following laboratory data were obtained through standardized assays and were collected from patients within 1 year of the baseline visit: serum creatinine, HDL-c, and HbA1c levels. QRS duration was recorded using MUSE electrocardiogram analysis software interfaced into the EHR database.

Re-derivation of the WATCH-DM risk score

Due to the lack of fasting plasma glucose and ECG variables in the validation cohorts, the WATCH-DM score was re-derived similar to the original score. Derivation was performed using data from the ACCORD and ALLHAT trials.[14, 15] Specifically, a Cox model was constructed with the same variables as the original WATCH-DM score and substituting fasting plasma glucose for HbA1c and removing QRS duration. An age-standardized points scoring system was then constructed using the methods described in the Framingham framework.[46] Models were internally trained and tested using a 70/30% derivation/validation cohort split. Discrimination and calibration performance of the rederived models is shown in **Figure S2**.

Implementation of a risk prediction API

To improve the clinical utility of our study, we implemented the ML model on a publicly available website (www.cvriskscores.com). The software components were designed to be used by any open source programming language through the use of a standardized programming language (or application programming interface [API]). Specifically, a REST API was created using the *plumber* package in R to send and receive data requests.[47, 48] Data request objects can be sent using the JavaScript Object Notation (JSON) format, commonly available in most programming languages.[49] To demonstrate the application, an example JSON object format to request the ML model is described below. The ML model was rederived using participant-level data from the ACCORD and ALLHAT trials. The HF risk prediction API is publicly accessible at watchdm.cvriskscores.com. The API accepts JSON requests and returns a JSON with the predicted risk and can accept multiple patient input. An example JSON request is shown below.

```
{
        "Age": 57,
        "BMI": 30,
        "SBP": 140,
        "DBP": 85,
        "SCreat": 1.3,
        "HbA1c": 7.8,
        “FPG”: 140,
        "HDL": 40,
        "QRS": 1,
        "hxMI": 0,
        "hxCABG":0,
}
```

Note, only HbA1c or FPG is required. QRS is a binary 0/1 indicating if the QRS duration is ≥ 120 ms. Finally, hxMI and hxCABG are binary 0/1 indicating the absence or presence of MI or CABG history, respectively.

Table S1. SNOMED CT heart failure intentional value set concepts and codes.

Concept	**SNOMED Code**
Congestive heart failure	42343007
Hypertensive heart failure	46113002
Diastolic heart failure	418304008
Systolic heart failure	417996009
Heart failure with normal ejection fraction	446221000
Left heart failure	85232009

Table S2. SNOMED CT heart failure extensional value set.

Concept ID	Preferred Term
364006	Acute left-sided heart failure
441481004	Chronic systolic heart failure
698594003	Symptomatic congestive heart failure
153951000119103	Acute on chronic combined systolic and diastolic heart failure
67431000119105	Congestive heart failure stage D
426611007	Congestive heart failure due to valvular disease
67441000119101	Congestive heart failure stage C
80479009	Acute right-sided congestive heart failure
194767001	Benign hypertensive heart disease with congestive cardiac failure
120861000119102	Systolic heart failure stage C
285211000119102	Congestive heart failure as post-operative complication of cardiac surgery
120851000119104	Systolic heart failure stage D
88805009	Chronic congestive heart failure
418304008	Diastolic heart failure
443253003	Acute on chronic systolic heart failure
10633002	Acute congestive heart failure
5148006	Hypertensive heart disease with congestive heart failure
120871000119108	Systolic heart failure stage B
195114002	Acute left ventricular failure
74960003	Acute left-sided congestive heart failure
15781000119107	Hypertensive heart AND chronic kidney disease with congestive heart failure
77737007	Benign hypertensive heart disease with congestive heart failure
46113002	Hypertensive heart failure
441530006	Chronic diastolic heart failure
23341000119109	Congestive heart failure with right heart failure
285221000119109	Congestive heart failure as post-operative complication of non-cardiac surgery
85232009	Left heart failure
698296002	Acute exacerbation of chronic congestive heart failure
120891000119109	Diastolic heart failure stage C
72481000119103	Congestive heart failure as early postoperative complication
153931000119109	Acute combined systolic and diastolic heart failure
92506005	Biventricular congestive heart failure
96311000119109	Exacerbation of congestive heart failure
111283005	Chronic left-sided heart failure
194781004	Hypertensive heart and renal disease with both (congestive) heart failure and renal failure
101281000119107	Congestive heart failure due to cardiomyopathy
443254009	Acute systolic heart failure
120901000119108	Diastolic heart failure stage B

67451000119104	Congestive heart failure stage B
83105008	Malignant hypertensive heart disease with congestive heart failure
194779001	Hypertensive heart and renal disease with (congestive) heart failure
43736008	Rheumatic left ventricular failure
42343007	Congestive heart failure
277638005	Sepsis-associated left ventricular failure
120881000119106	Diastolic heart failure stage D
446221000	Heart failure with normal ejection fraction
443344007	Acute on chronic diastolic heart failure
82523003	Congestive rheumatic heart failure
443343001	Acute diastolic heart failure
66989003	Chronic right-sided congestive heart failure
417996009	Systolic heart failure
153941000119100	Chronic combined systolic and diastolic heart failure
426263006	Congestive heart failure due to left ventricular systolic dysfunction
5375005	Chronic left-sided congestive heart failure
71892000	Cardiac asthma

Table S3. Baseline characteristics of participants in the Look AHEAD cohort stratified by quintiles of WATCH-DM(i) scores.

	WATCH-DM ≤ 11	**WATCH-DM 12**	**WATCH-DM 13-14**	**WATCH-DM 15-16**	**WATCH-DM ≥ 17**
N	1493	573	1265	841	695
Age, years	56 [51, 59]	58.0 [54, 63]	59 [55, 64]	61 [57, 66]	64 [59, 68]
Men	539 (36.1)	222 (38.7)	526 (41.6)	366 (43.5)	359 (51.7)
White race	906 (60.7)	371 (64.7)	844 (66.7)	593 (70.5)	514 (74.0)
Body mass index, kg/m^2	32.3 [29.8, 34.8]	33.9 [31.0, 38.0]	35.9 [32.4, 40.2]	37.3 [33.8, 41.6]	38.4 [35.2, 42.7]
Systolic BP, mm Hg	124 [115, 134]	125 [116, 137]	128 [116, 140]	133 [119, 145]	138 [119, 149]
Diastolic BP, mm Hg	72 [66, 79]	70 [65, 76]	69 [63, 76]	69 [61, 75]	68 [59, 74]
Diabetes duration, years	4.0 [2.0, 7.0]	5.0 [2.0, 8.0]	5.0 [2.0, 10.0]	6.0 [3.0, 10.0]	7.0 [3.5, 12.0]
Serum creatinine, mg/dL	0.8 [0.7, 0.9]	0.8 [0.7, 0.9]	0.8 [0.7, 1.0]	0.8 [0.7, 1.0]	0.9 [0.7, 1.0]
HbA1c, %	6.7 [6.2, 7.4]	7.0 [6.5, 7.7]	7.0 [6.5, 7.8]	7.3 [6.7, 8.3]	7.7 [7.0, 9.1]
HDL-c, mg/dL	44 [37, 57]	42.0 [37, 49]	41 [35, 49]	41 [34, 48]	38 [32, 47]
Urine albumin-creatinine ratio, mg/g	7.2 [4.8, 14.0]	7.9 [4.9, 18.4]	8.9 [5.4, 21.2]	10.3 [6.1, 23.2]	13.9 [6.6, 43.8]
Prior MI	3 (0.2)	7 (1.2)	32 (2.5)	52 (6.2)	193 (27.8)
Prior CABG	3 (0.2)	0 (0.0)	7 (0.6)	29 (3.4)	80 (11.5)
Insulin use	182 (12.2)	98 (17.1)	229 (18.1)	183 (21.8)	204 (29.4)

Values are displayed as median (25^{th}, 75^{th} percentiles) for continuous and number (percent) for categorical variables.

Abbreviations: BP, blood pressure; CABG, coronary artery bypass graft; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; Look AHEAD, Look Action for Health in Diabetes; MI, myocardial infarction; WATCH-DM(i), integer-based WATCH-DM model.

Table S4. Number of HF events, Kaplan-Meier HF risk estimate, number of participants across WATCH-DM and TRS-HF$_{DM}$ strata and cohorts. The 5-year risk of incident heart failure was estimated in the Look AHEAD and EHR cohorts while the 4-year risk, due to a shorter duration of follow-up, was estimated in the TECOS cohort.

	Look AHEAD			**TECOS**			**EHR**		
	Num. HF events	**KM risk estimate (95% CI)**	**Participants**	**Num. HF events**	**KM risk estimate (95% CI)**	**Participants**	**Num. HF events**	**KM risk estimate (95% CI)**	**Participants**
Overall	91	1.83 (1.44, 2.21)	4867	266	3.11 (2.71, 3.58)	12028	133	1.88 (1.56, 2.21)	7475
WATCH-DM									
Quintile 1	13	0.90 (0.41, 1.39)	1493	31	1.85 (1.24, 2.76)	2447	9	0.56 (0.19, 0.93)	1836
Quintile 2	6	1.00 (0.43, 1.53)	573	37	2.14 (1.48, 3.09)	2358	13	0.96 (0.44, 1.47)	1499
Quintile 3	16	1.81 (0.69, 2.91)	1265	42	2.21 (1.58, 3.10)	2487	11	1.65 (0.96, 2.34)	1489
Quintile 4	16	1.98 (1.01, 2.93)	841	36	2.52 (1.65, 3.83)	2035	38	2.73 (1.83, 3.61)	1522
Quintile 5	40	5.97 (4.16, 7.75)	695	120	6.66 (5.39, 8.22)	2701	62	5.29 (3.87, 6.68)	1129
TRS-HF$_{DM}$									
0	29	0.91 (0.58, 1.24)	3291	6	1.16 (0.41, 3.27)	675	-	-	-
1	35	2.99 (2.01, 3.96)	1222	39	2.55 (1.79, 3.64)	2206	-	-	-
2	18	6.24 (3.41, 8.99)	297	33	4.07 (2.74, 6.03)	1127	-	-	-
3+	9	16.10 (5.89, 25.30)	57	21	6.42 (4.13, 9.90)	400	-	-	-

Abbreviations: CI, confidence interval; EHR, electronic health record; HF, heart failure; KM, Kaplan-Meier; Look AHEAD, Look Action for Health in Diabetes; TECOS, Trial Evaluating Cardiovascular Outcomes With Sitagliptin.

Table S5. Association of WATCH-DM integer-based risk score quintiles with risk of incident heart failure (HF) in the Look AHEAD, TECOS, and electronic health record cohorts.

WATCH-DM(i)		***TRS-HF$_{DM}$***	
Score Quintile	Hazard ratio (95% CI)	Score Category	Hazard ratio (95% CI)
Look AHEAD			
Quintile 1	Ref.	0	Ref.
Quintile 2	1.33 (0.46–3.84)	1	3.32 (2.03-5.43)
Quintile 3	2.52 (1.02–6.49)	2	6.98 (3.88-12.58)
Quintile 4	2.86 (1.13–7.26)	3+	20.04 (9.49-42.34)
Quintile 5	6.79 (3.32–15.27)		
TECOS			
Quintile 1	Ref.	0	Ref.
Quintile 2	1.27 (0.79–2.04)	1	1.99 (0.84, 4.71
Quintile 3	1.38 (0.87–2.19)	2	3.47 (1.46, 8.29)
Quintile 4	1.43 (0.89–2.32)	3+	6.71 (2.71, 16.60)
Quintile 5	3.78 (2.54–5.61)		
Electronic Health Record			
Quintile 1	Ref.	-	-
Quintile 2	1.77 (0.76-4.15)	-	-
Quintile 3	3.00 (1.24-7.24)	-	-
Quintile 4	4.35 (2.10-9.00)	-	-
Quintile 5	8.98 (4.46-18.06)	-	-

Abbreviations: CI, confidence interval; Look AHEAD, Look Action for Health in Diabetes; TECOS, Trial Evaluating Cardiovascular Outcomes With Sitagliptin.

Table S6. Baseline characteristics of participants in the Look AHEAD cohort stratified by TRS-HF$_{DM}$ score categories.

	TRS-HF$_{DM}$ = 0	**TRS-HF$_{DM}$ = 1**	**TRS-HF$_{DM}$ = 2**	**TRS-HF$_{DM}$ = 3+**
N	3291	1222	297	57
Age, years	58 [54, 62]	60 [55, 65]	62 [58, 67]	64 [59, 69]
Men	1224 (37.2)	572 (46.8)	179 (60.3)	37 (64.9)
White race	2130 (64.7)	844 (69.1)	214 (72.1)	40 (70.2)
Body mass index, kg/m^2	34.9 [31.5, 39.2]	35.0 [31.6, 39.5]	35.5 [32.3, 40.5]	35.4 [31.6, 41.8]
Systolic BP, mm Hg	127 [116, 139]	130 [117, 142]	135 [119, 149]	134 [121, 142]
Diastolic BP, mm Hg	70.5 [63, 77]	70 [63, 76]	71 [63, 77]	70 [63, 77]
Diabetes duration, years	5.0 [2.0, 8.0]	6.0 [3.0, 10.0]	7.0 [4.0, 15.0]	9.0 [4.0, 15.0]
Serum creatinine, mg/dL	0.8 [0.7, 0.9]	0.9 [0.7, 1.0]	1.0 [0.8, 1.1]	1.2 [0.9, 1.3]
HbA1c, %	7.0 [6.4, 7.8]	7.2 [6.5, 8.1]	7.2 [6.7, 8.0]	7.4 [6.9, 8.2]
HDL-c, mg/dL	43 [36, 51]	41 [34, 48]	39 [33, 47]	40 [35, 51]
Urine albumin-creatinine ratio, mg/g	7.2 [4.9, 11.8]	20.5 [7.0, 51.2]	79.1 [37.8, 377.1]	401.1 [65.8, 696.9]
Prior MI	0 (0.0)	191 (15.6)	81 (27.3)	15 (26.3)
Prior CABG	0 (0.0)	71 (5.8)	33 (11.1)	15 (26.3)
Insulin use	505 (15.3)	273 (22.3)	98 (33.0)	20 (35.1)

Values are displayed as median (25th, 75th percentiles) for continuous and number (percent) for categorical variables.

Abbreviations: BP, blood pressure; CABG, coronary artery bypass graft; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; Look AHEAD, Look Action for Health in Diabetes; MI, myocardial infarction; WATCH-DM(i), integer-based WATCH-DM model.

Table S7. Baseline characteristics of participants in the TECOS cohort stratified by quintiles of WATCH-DM(i) scores.

	WATCH-DM ≤ 11	**WATCH-DM 12 - 13**	**WATCH-DM 14 - 15**	**WATCH-DM 16 - 17**	**WATCH-DM ≥ 18**
N	2,447	2,358	2,487	2,035	2,701
Age, years	59 (55-64)	63 (58-68)	65 (60-70)	67 (62-72)	70 (65-75)
Men	62.9% (1,539)	68.8% (1,623)	71.4% (1,776)	76.9% (1,565)	80.2% (2,165)
White race	50.7% (1,240)	59.3% (1,399)	64.6% (1,606)	68.7% (1,399)	78.4% (2,117)
Body mass index, kg/m^2	27.3 (24.8-30.1)	28.4 (25.5-31.6)	29.2 (25.9-32.7)	30.1 (26.8-34.2)	31.5 (27.8-36.0)
Systolic BP, mm Hg	130 (122-138)	130 (122-140)	135 (124-147)	135 (124-147)	140 (124-150)
Diastolic BP, mm Hg	80 (76-85)	80 (70-85)	78 (70-84)	75 (69-81)	72 (66-80)
Diabetes duration, years	8 (4-13)	9 (5-14)	10 (6-16)	11 (6-16)	12 (7-19)
Serum creatinine, mg/dL	0.9 (0.8-1.0)	0.9 (0.8-1.1)	1.0 (0.8-1.1)	1.0 (0.9-1.2)	1.1 (0.9-1.3)
HDL-c, mg/dL	46 (39-59)	42 (36-50)	41 (35-49)	41 (35-47)	39 (33-46)
HbA1c, %	7.1 (6.7-7.6)	7.2 (6.8-7.6)	7.2 (6.8-7.7)	7.3 (6.8-7.7)	7.3 (6.9-7.7)
Prior MI	10.0% (245)	26.2% (617)	36.9% (918)	51.3% (1,043)	68.3% (1,844)
Prior CABG	3.1% (77)	10.0% (235)	18.5% (461)	30.4% (618)	57.6% (1,555)
Cerebrovascular disease	34.0% (833)	28.2% (665)	24.3% (604)	18.3% (372)	16.5% (446)
Peripheral arterial disease	26.3% (644)	18.3% (431)	16.2% (402)	13.5% (274)	11.9% (322)
Insulin use	14.1% (344)	17.5% (413)	20.9% (520)	23.8% (485)	32.5% (877)

Values are displayed as median (25^{th}, 75^{th} percentiles) for continuous and number (percent) for categorical variables.

Abbreviations: BP, blood pressure; CABG, coronary artery bypass graft; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; MI, myocardial infarction; TECOS, Trial Evaluating Cardiovascular Outcomes with Sitagliptin; WATCH-DM(i), integer-based WATCH-DM model.

Table S8. Baseline characteristics of participants in the TECOS cohort stratified by TRS-HF_{DM} score categories.

	TRS-HF_{DM} = 0	**TRS-HF_{DM} = 1**	**TRS-HF_{DM} = 2**	**TRS-HF_{DM} = 3+**
N	675	2206	1127	400
Age, years	63 [57-69]	65 [59-70]	67 [62-73]	69 [64-75]
Men	412 (61.0)	1165 (75.5)	888 (78.8)	309 (77.3)
White race	371 (55.0)	1499 (68.0)	809 (71.8)	294 (73.5)
Body mass index, kg/m^2	28.4 [24.9-32.1]	29.0 [26.0-32.7]	29.8 [26.7-33.3]	30.1 [26.7-34.4]
Systolic BP, mm Hg	134 [128-143]	131 [122-142]	135 [124-146]	135 [125-146]
Diastolic BP, mm Hg	80 [71-87]	77 [69-82]	76 [68-82]	72 [65-80]
Serum creatinine, mg/dL	0.9 [0.8-1.1]	0.9 [0.8-1.1]	1.1 [0.9-1.3]	1.3 [1.0-1.5]
HDL-c, mg/dL	7.2 [6.8-7.6]	7.2 [6.8-7.6]	7.2 [6.8-7.7]	7.3 [6.9-7.7]
HbA1c, %	44 [38-50]	42 [35-50]	41 [34-49]	40 [34-47]
Prior MI	0 (0.0)	1021 (46.3)	569 (50.5)	214 (53.5)
Prior CABG	0 (0.0)	601 (27.2)	335 (31.5)	147 (36.8)
Cerebrovascular disease	51.6% (348)	17.0% (374)	17.3% (195)	19.5% (78)
Peripheral arterial disease	50.7% (342)	12.7% (281)	11.4% (129)	14.3% (57)
Insulin use	98 (14.5)	413 (18.7)	284 (25.2)	145 (36.3)

Values are displayed as median (25^{th}, 75^{th} percentiles) for continuous and number (percent) for categorical variables.
Abbreviations: BP, blood pressure; CABG, coronary artery bypass graft; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; MI, myocardial infarction; TECOS, Trial Evaluating Cardiovascular Outcomes with Sitagliptin.

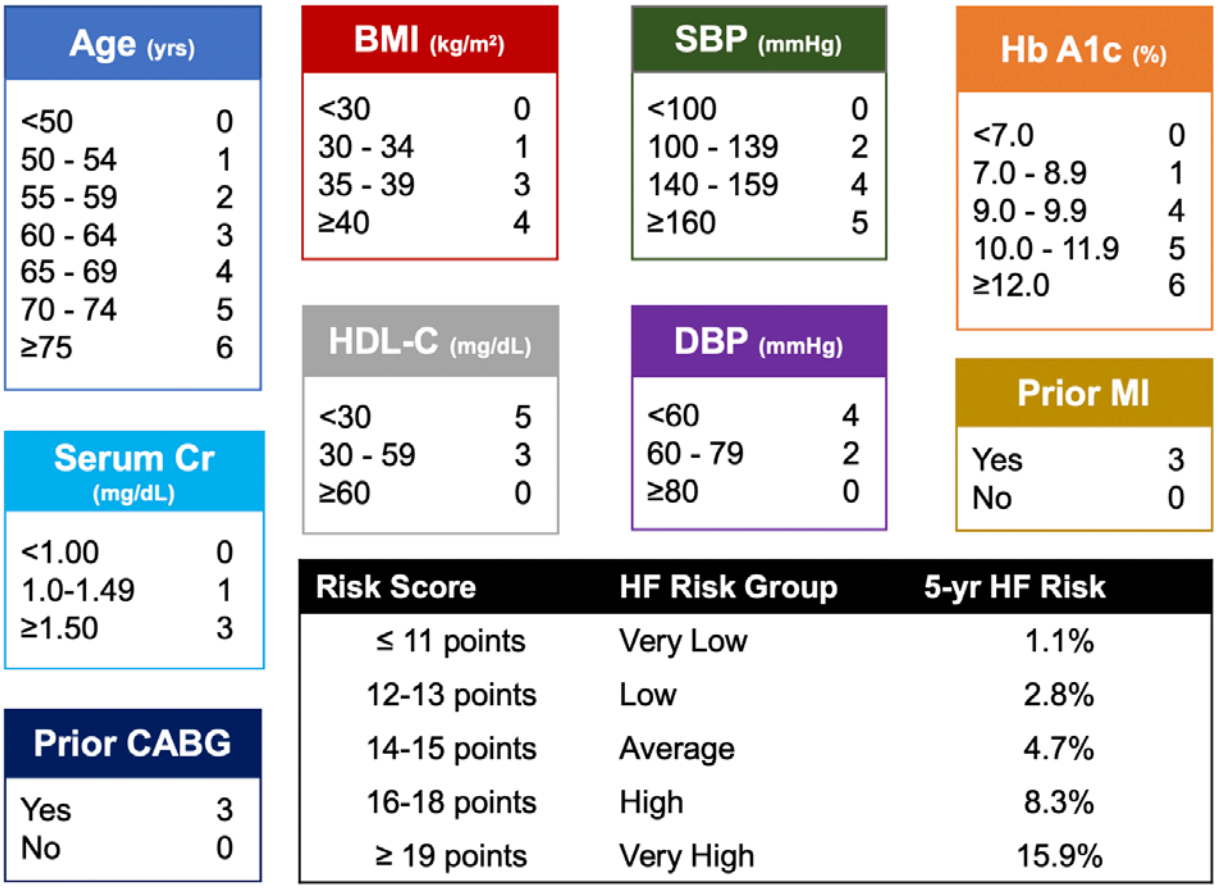
Table S9. Baseline characteristics of participants in the EHR cohort stratified by quintiles of WATCH-DM(i) scores.

	WATCH-DM ≤ 7	**WATCH-DM 8-9**	**WATCH-DM 10-11**	**WATCH-DM 12-14**	**WATCH-DM ≥ 15**
N	1836	1499	1489	1522	1129
Age, years	50 [41, 57]	59 [49, 67]	63 [54, 71]	65 [58, 72]	68 [61, 74]
Men	797 (43.4)	669 (44.6)	771 (51.8)	823 (54.1)	664 (58.8)
White race	945 (51.5)	816 (54.4)	839 (56.3)	826 (54.3)	657 (58.2)
Body mass index, kg/m^2	29.0 [24.9, 33.4]	30.4 [26.0, 35.7]	31.0 [27.3, 36.2]	32.0 [27.5, 38.7]	33.5 [28.4, 40.8]
Systolic BP, mm Hg	126 [116, 136]	130 [120, 143]	134 [121, 148]	135 [122, 150]	137 [122, 153]
Diastolic BP, mm Hg	78 [71, 84]	78 [70, 83]	75 [68, 83]	74 [66, 82]	72 [65, 78]
Serum creatinine, mg/dL	0.8 [0.6, 0.9]	0.8 [0.7, 1.0]	0.9 [0.8, 1.2]	1.1 [0.8, 1.4]	1.5 [1.1, 2.1]
HDL-c, mg/dL	48 [44, 56]	48 [44, 52]	47 [41, 50]	46 [41, 49]	45 [37, 48]
HbA1c, %	7.0 [6.3, 7.8]	7.3 [6.4, 7.9]	7.3 [6.6, 7.9]	7.3 [6.7, 8.1]	7.4 [7.0, 9.1]
Prior MI	11 (0.6)	54 (3.6)	138 (9.3)	289 (19.0)	500 (44.3)
Prior CABG	11 (0.6)	34 (2.3)	69 (4.6)	169 (11.1)	309 (27.4)

Values are displayed as median (25^{th}, 75^{th} percentiles) for continuous and number (percent) for categorical variables.

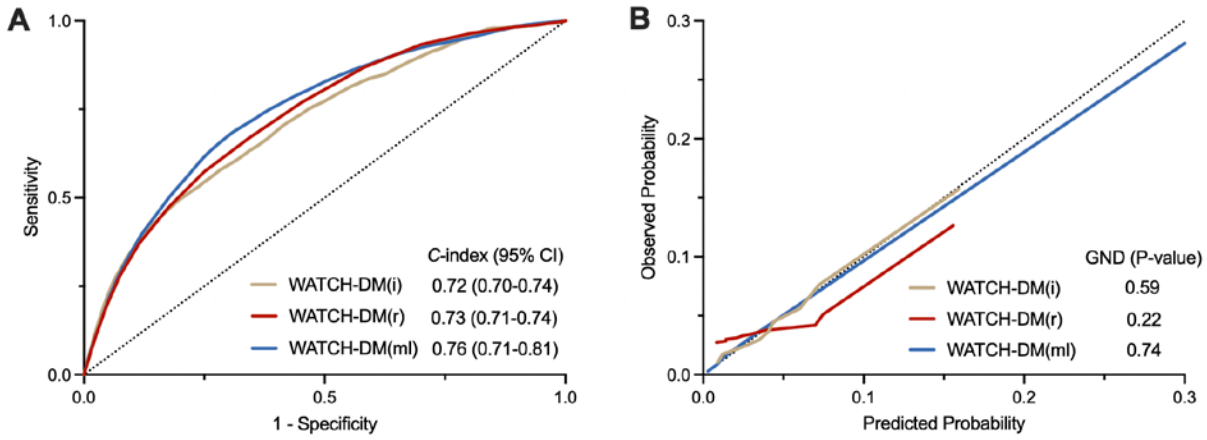
Abbreviations: BP, blood pressure; CABG, coronary artery bypass graft; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; MI, myocardial infarction; WATCH-DM(i), integer-based WATCH-DM model.

Figure S1. The re-derived WATCH-DM integer-based risk score after substituting hemoglobin A1c for fasting plasma glucose and excluding electrocardiographic parameters.



Abbreviations: BMI, body mass index; CABG, coronary artery bypass graft; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-c, high-density lipoprotein cholesterol; HF, heart failure; MI, myocardial infarction; SBP, systolic blood pressure.

Figure S2. A) Discrimination (receiver operator characteristics) and **B)** calibration plots of the re-derived WATCH-DM risk scores in the derivation (ACCORD and ALLHAT) cohort. Models were trained and tested using a 70/30% derivation/validation cohort split.



Abbreviations: ACCORD, Action to Control Cardiovascular Risk in Diabetes; ALLHAT, Antihypertensive and Lipid-Lowering Treatment Prevent Heart Attack Trial; CI, confidence interval; GND, Greenwood-Nam-D'Agostino statistic; WATCH-DM(i), integer-based WATCH-DM model; WATCH-DM(r), regression-based WATCH-DM model; WATCH-DM(ml), machine learning-based WATCH-DM model.

Figure S3. Calibration plots of the **A)** integer-based WATCH-DM(i), **B)** regression-based WATCH-DM(r), and **C)** machine learning-based WATCH-DM(ml) risk scores in the Look AHEAD, TECOS, and electronic health record (EHR) external validation cohorts.

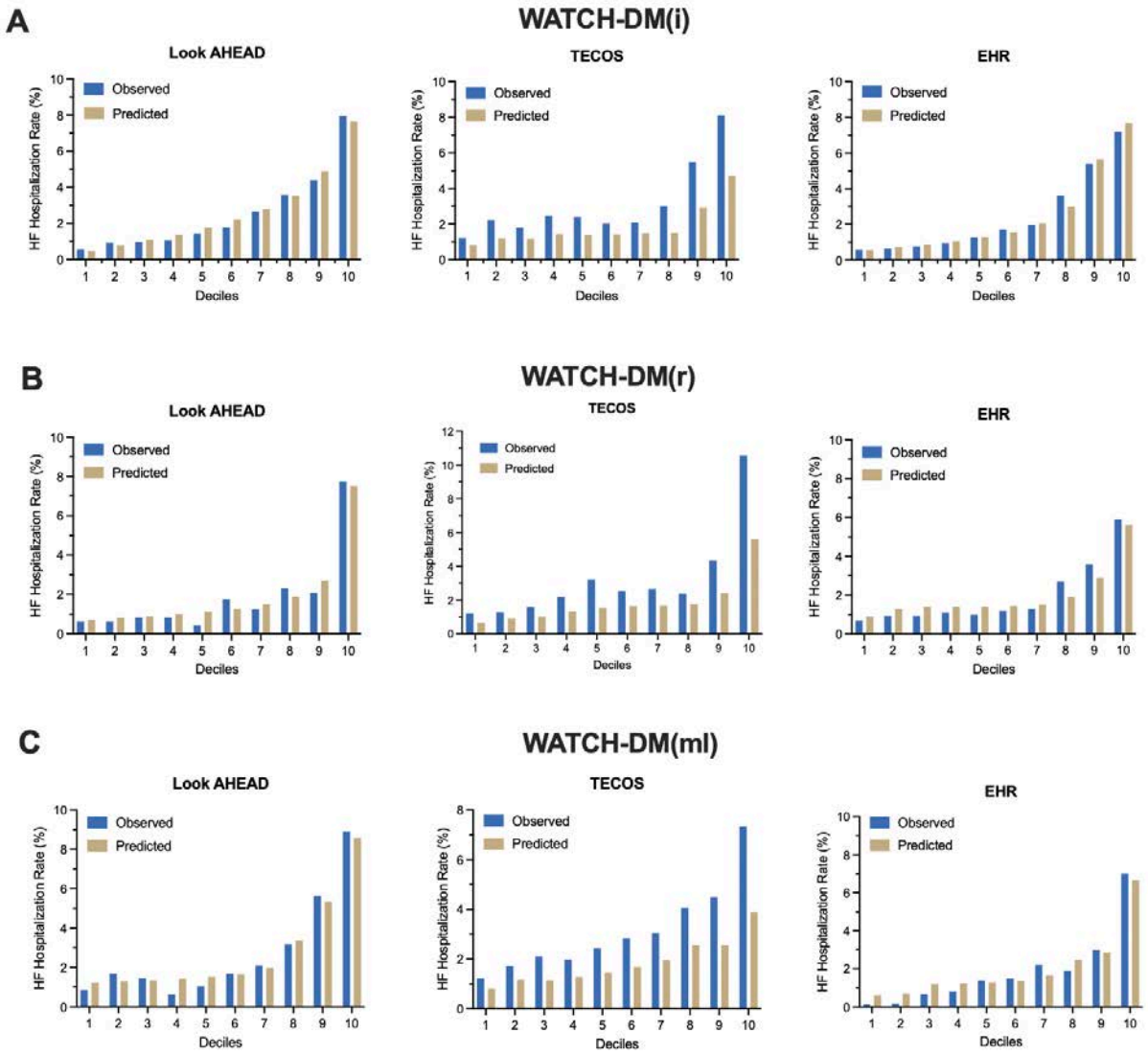


Figure S4. Calibration plots of the integer-based WATCH-DM(i) score categories in the **A)** Look AHEAD, **B)** TECOS, and **C)** electronic health record (EHR) external validation cohorts.

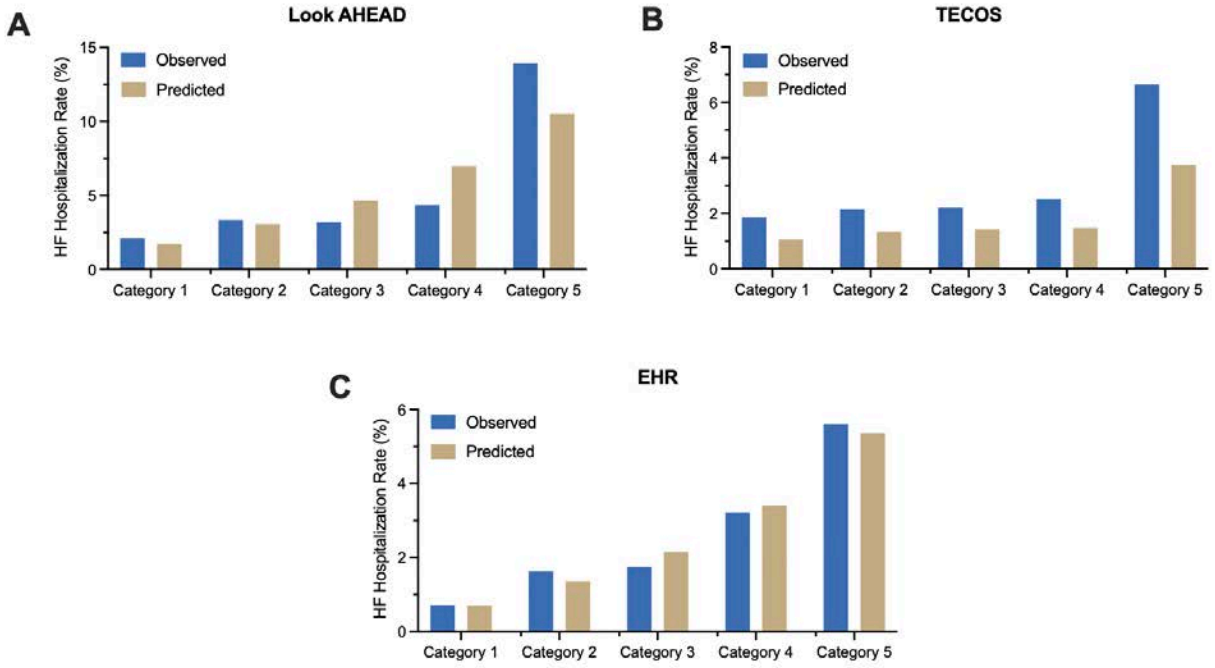


Figure S5. Calibration plots of the TRS-HF$_{DM}$ risk score in the **A)** Look AHEAD and **B)** TECOS external validation cohorts.