


Research and Applications

Cardiovascular disease risk prediction for people with type 2 diabetes in a population-based cohort and in electronic health record data

Jackie Szymonifka ¹, Sarah Conderino,² Christine Cigolle,^{3,4,5} Jinkyung Ha,^{3,4,5} Mohammed Kabeto,^{3,4,5} Jaehong Yu,¹ John A. Dodson,^{6,7} Lorna Thorpe,² Caroline Blaum⁸, and Judy Zhong¹

¹Division of Biostatistics, Department of Population Health, NYU Langone Health, New York, New York, USA, ²Division of Epidemiology, Department of Population Health, NYU Langone Health, New York, New York, USA, ³Department of Family Medicine, University of Michigan, Ann Arbor, Michigan, USA, ⁴Geriatric Research, Education and Clinical Center (GRECC), VA Ann Arbor Healthcare System, Ann Arbor, Michigan, USA, ⁵Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA, ⁶Leon H. Charney Division of Cardiology, Department of Medicine, New York University School of Medicine, New York, New York, USA, ⁷Division of Healthcare Delivery Science, Department of Population Health, New York University School of Medicine, New York, New York, USA and ⁸Division of Geriatric Medicine and Palliative Care, Department of Medicine, New York University School of Medicine, New York, New York, USA

Corresponding Author: Judy Zhong, Associate Professor and Director, Biostatistics, Epidemiology and Research Design Program (BERD), NYU-H+H Clinical and Translational Science Institute, NYU Langone Health, Division of Biostatistics, Department of Population Health, 180 Madison Avenue, 4th Floor, Room 452, New York, NY 10016, USA (judy.zhong@nyumc.org)

Received 2 September 2020; Revised 13 October 2020; Accepted 20 October 2020

ABSTRACT

Objective: Electronic health records (EHRs) have become a common data source for clinical risk prediction, offering large sample sizes and frequently sampled metrics. There may be notable differences between hospital-based EHR and traditional cohort samples: EHR data often are not population-representative random samples, even for particular diseases, as they tend to be sicker with higher healthcare utilization, while cohort studies often sample healthier subjects who typically are more likely to participate. We investigate heterogeneities between EHR- and cohort-based inferences including incidence rates, risk factor identifications/quantifications, and absolute risks.

Materials and methods: This is a retrospective cohort study of older patients with type 2 diabetes using EHR from New York University Langone Health ambulatory care (NYULH-EHR, years 2009–2017) and from the Health and Retirement Survey (HRS, 1995–2014) to study subsequent cardiovascular disease (CVD) risks. We used the same eligibility criteria, outcome definitions, and demographic covariates/biomarkers in both datasets. We compared subsequent CVD incidence rates, hazard ratios (HRs) of risk factors, and discrimination/calibration performances of CVD risk scores.

Results: The estimated subsequent total CVD incidence rate was 37.5 and 90.6 per 1000 person-years since T2DM onset in HRS and NYULH-EHR respectively. HR estimates were comparable between the datasets for most demographic covariates/biomarkers. Common CVD risk scores underestimated observed total CVD risks in NYULH-EHR.

Discussion and conclusion: EHR-estimated HRs of demographic and major clinical risk factors for CVD were mostly consistent with the estimates from a national cohort, despite high incidences and absolute risks of total CVD outcome in the EHR samples.

LAY SUMMARY

As we transition from paper charts to electronic health records (EHRs), there is an unprecedented opportunity to use all available data within the EHR for public health research. Despite this advantage, it is important to recognize that EHR often are not random samples from the population and risk prediction models generated from EHR may be biased. In this study, we calculated the risk of subsequent cardiovascular disease following type 2 diabetes mellitus diagnosis in two cohorts: the Health and Retirement Study (HRS), which is a nationally representative longitudinal study, and the NYU Langone Health ambulatory care EHR. Although identical eligibility criteria were used for both cohorts, EHR patients tended to be sicker than their HRS counterparts. Despite these differences, risk estimates were generally comparable between the two samples. This is encouraging as it provides credence that EHR may be used to generate risk prediction estimates of the general population.

Key words: electronic health records, cohort analysis, risk factors, cardiovascular disease, type 2 diabetes mellitus

BACKGROUND AND SIGNIFICANCE

As healthcare systems transition from paper charts to electronic health records (EHRs), investigators have an unprecedented opportunity to use these rich data sources for public health research purposes on chronic diseases.¹⁻³ It is important to acknowledge that there may be notable differences between hospital-based EHR and traditional cohort samples, the latter of which often have representative sampling weights, regular sampling time and follow-up, and standardized data collection metrics.⁴ With the exception of universal-registration EHR such as in the United Kingdom or Scottish Care Information-Diabetes (SCI-Diabetes database), EHR often are not population-representative random samples, even for populations with a particular disease, as EHR samples tend to be sicker and have higher healthcare utilization.^{5,6} They often are limited to patients presenting to a particular healthcare provider, which may further limit their representation of the general population.⁷ In contrast, cohort studies often sample healthier participants who typically are more likely to respond to surveys.⁸

Type 2 diabetes mellitus (T2DM) has become one of the most prevalent chronic diseases worldwide, estimated to have affected 8.8% of the population in 2017,⁹ and this prevalence is expected to continue rising.¹⁰ The International Diabetes Federation estimated the 2015 direct global cost of diabetes at \$673 billion and predicted this will rise to approximately \$802 billion by 2040.¹¹⁻¹⁴ Cardiovascular disease (CVD) is the most prevalent cause of morbidity and mortality among T2DM patients, affecting approximately 32% and causing 47% of deaths within this patient population.¹⁵⁻¹⁷ As the population ages and T2DM prevalence increases, improving our understanding of CVD heterogeneity is key to minimizing long-term complications in elders.¹⁸

Traditionally, CVD risk prediction algorithms are developed utilizing data from clinical trials and cohort studies.¹⁹⁻²⁴ Examples include the original Framingham Risk Score (FRS) to predict hard CVD risk,²¹ subsequently updated to provide total CVD risk estimate.²³ The American College of Cardiology/American Heart Association (ACC/AHA) pooled cohort atherosclerotic CVD risk equations were developed to predict the risk of hard CVD events based on the same factors as the FRS.¹⁹ There also have been T2DM-specific total and hard CVD risk predictions, including QRISK2,²⁴ ADVANCE,²⁵ and Swedish National Diabetes Register (NDR)²⁶ risk scores. Internal c-statistics for these risk scores ranged

between 0.64 and 0.82, but an external study yielded c-statistics of 0.66–0.67 for each of these scores, with the weakest discrimination among older adults.²⁷

OBJECTIVE

The purpose of this study is to investigate the heterogeneities between EHR- and cohort-based inferences including incidence rates (IRs), identifications and quantifications of risk factors, and absolute risks (ARs). We investigate these heterogeneities by comparing risk predictions of the same outcome (subsequent incident CVD) in the same target population (incident T2DM adult patients) using the same eligibility criteria in a nationwide U.S. cohort, the Health and Retirement Survey (HRS),²⁸ and a large hospital EHR-based ambulatory care sample from New York University Langone Health (NYULH-EHR). We hypothesized that EHR estimates of CVD risk factors would be population-generalizable despite higher CVD IRs and that AR predictions derived from population-based cohorts would tend to underestimate actual risk observed in EHR samples.

MATERIALS AND METHODS

Study design and participants

We define inclusion criteria for both samples as age ≥ 50 years, with incident T2DM and no CVD (defined below) at T2DM incidence. We chose incident T2DM since the year of T2DM diagnosis can be consistently defined without recall bias.

Within NYULH-EHR, eligible patients are defined as having a visit history of ≥ 1 year with an NYULH ambulatory primary care physician between 2009 and 2017 and having T2DM by a rule-based phenotyping algorithm. We excluded patients seen for consultation only and patients in the emergency department, inpatient, or specialist settings, as these lacked consistent T2DM documentation. We defined T2DM incidence as any of (1) ≥ 2 encounters with a T2DM *International Classification of Diseases* (ICD)-10 code, or (2) ≥ 2 abnormal HbA1C results ($\geq 6.5\%$) and ≥ 1 encounter with a T2DM ICD-10 code, or (3) a T2DM medication prescription, excluding metformin and acarbose. The date of T2DM is defined as the date the first T2DM diagnostic code was recorded. Among NYULH-EHR patients, we excluded patients with T2DM diagnosis at EHR-entry as the diagnosis date was unknown. We excluded

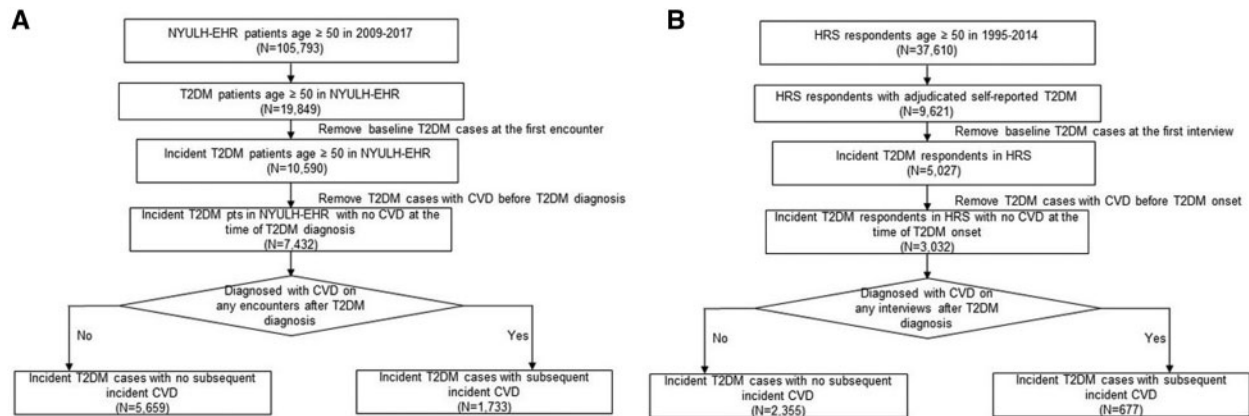


Figure 1. Inclusion criteria and study flow chart for (A) NYULH-EHR T2DM patients and (B) HRS T2DM respondents. (A) Inclusion criteria for the NYULH-EHR cohort. Among patients seen in the New York University Langone Health ambulatory care clinic between 1995 and 2014 who met the eligibility criteria outlined in [Supplementary Figure S1](#), we first limited the analysis cohort to encounters with patients 50 years of age or older. We then reduced the eligible cohort to patients ≥ 50 years of age who had T2DM, as defined in [Supplementary Figure S1](#). We removed patients who met the criteria for T2DM status at initial encounter since date of initial diagnosis could not be reliably estimated. Finally, we removed patients who already met the criteria for CVD diagnosis, so that subsequent incident CVD cases could be identified. (B) Inclusion criteria for the HRS cohort. Among respondents to the HRS between 1992 and 2014, we first limited the analysis cohort to respondents who were 50 years of age or older during at least one interview. We then reduced the eligible cohort to respondents age ≥ 50 years with self-reported and subsequently adjudicated T2DM.³⁴ We also eliminated T2DM cases that were self-reported at the first interview since date of initial diagnosis could not be reliably estimated. Finally, we removed respondents with self-reported, and subsequently adjudicated,³⁴ CVD or stroke at or before the first interview at which T2DM was reported so that subsequent incident CVD cases could be identified.

T2DM patients with an ICD-10 code for CVD (defined below) preceding or at T2DM diagnosis ([Figure 1A](#)).

HRS is a biennial longitudinal health interview survey of over 37 620 older adults in the United States and is well-established as nationally representative.^{28–32} A multistage, national area-clustered probability sampling frame was used and blacks, Hispanics, and Florida residents were oversampled to ensure adequate representation.³³ Baseline response rates averaged 73.0%, while follow-up response rates were 68.8–92.3%.³³ Each HRS biennial wave provides self-reported chronic diseases, including T2DM, and was adjudicated using HbA1C and medication information to address inconsistencies across waves.³⁴ Year of T2DM onset was estimated as the year of the first-reported T2DM without T2DM in previous wave(s). We excluded respondents who reported T2DM at entry, as diagnosis year was unknown, and respondents who reported CVD before or concurrently to T2DM onset ([Figure 1B](#)). This study was approved by the NYULH-IRB and data were deidentified to assure anonymity.

Outcome

Total CVD was defined in NYULH-EHR as ≥ 3 encounters with an ICD-10 code for ischemic heart disease, ischemic or hemorrhagic cerebrovascular events, peripheral vascular disease, congestive heart failure, or other ischemic CVD ([Supplementary Table S2](#)).^{23,35,36} Hard CVD was defined as ≥ 1 encounter with an ICD-10 code for myocardial infarction, ischemic stroke, hemorrhagic stroke, or hospitalization or death from other CVD outcomes ([Supplementary Table S2](#)).^{19,35,36} Date of CVD is defined as the encounter date when the first CVD diagnostic code was recorded. Time-to-CVD was defined as the time between T2DM diagnosis and CVD diagnosis for CVD cases. Patients without subsequent CVD were censored with time-to-event defined as the time from T2DM diagnosis to the last known encounter.

Total CVD was defined by self-reported heart disease or stroke in HRS. We adjudicated self-reported outcomes using the same adju-

ication methodology.³⁴ Year of incident CVD onset was estimated as the year when the respondent first reported heart disease or stroke. Time-to-CVD was defined as the time between the year of T2DM onset and the year of CVD for CVD cases. Patients without subsequent CVD were censored with time-to-event defined as the time from T2DM onset to the year of the last interview.

Risk factors

CVD risk factors considered include age, sex, race, ethnicity, smoking, and biomarkers, including systolic blood pressure (SBP), body mass index (BMI), HbA1c, total cholesterol (TC), and high-density lipoprotein (HDL). Smoking was defined as a previous or current smoker. Additional NYULH-EHR risk factors included albumin, creatinine, and estimated glomerular filtration rate (eGFR). Blood-based biomarker data, including TC, HDL, HbA1c, C-reactive protein (CRP), and cystatin C (CysC), between 2006 and 2014 were used in HRS. Biomarker collection is staggered such that alternate halves of the HRS sample are tested in each biennial core survey.^{37,38} NYULH-EHR medication prescriptions were grouped by therapeutic class codes of antihyperglycemic, antihypertensive, and lipid-lowering, and prescriptions any time during follow-up were considered the use of the respective medication. Self-report antihypertensive and antihyperglycemic medication was included in HRS biennial core surveys, and self-reports at any time were considered the use of the respective medication.²⁸

Absolute 5-year CVD risk scores

Three established CVD risk functions were evaluated: the FRS total CVD function²³ in both the NYULH-EHR and HRS cohorts, and the ACC/AHA¹⁹ and Swedish NDR²⁶ risk prediction functions for hard CVD outcomes in the NYULH-EHR cohort only. For FRS total CVD and ACC/AHA functions, age, TC, HDL, and SBP values were defined with measurements at T2DM onset, or at the closest measurements following T2DM onset if no concurrent measurements. Hypertension treatment was defined with antihypertensive

prescriptions occurring within 3 months preceding the predictor measurement date in NYULH-EHR, and with self-report antihyperglycemic usage on the year of predictor measurement in HRS. Sex-specific equations were used. For the ACC/AHA risk score, African American (AA) coefficients were used for AAs and white coefficients for all others. We replaced 10-year baseline survival estimates with 5-year estimates by assuming exponential survival distributions to align with available follow-up of our cohorts.³⁹ For the Swedish NDR risk prediction, measurements were defined as described above and T2DM duration was defined as years from T2DM diagnosis to the date of predictor measurements. Lipid-reducing treatment was defined by prescriptions occurring within the 3 months preceding the predictor measurement date.

Statistical analysis

Demographics and risk factors were summarized using means (standard deviations) for continuous measures and frequencies (percentages) for categorical. Comparisons were performed using two-sample *t*-tests and chi-squared tests, respectively. IRs of subsequent CVD were estimated from the time of T2DM onset. Time-to-CVD was analyzed using Cox proportional hazard models. Model 1 included demographic covariates only, including age at T2DM onset, sex, race, ethnicity, and smoking status; while models 2–11 included the demographic covariates from model 1 and a biomarker in each model. We included the annual number of encounters as an additional covariate in NYULH-EHR models to mitigate the bias in EHRs that patients with CVD are likely to have more encounters; and compared with the analyses without adjusting for encounters.⁵ Hazard ratios (HRs) (HRs) were estimated per 1-standard deviation increase in the respective biomarker. When multiple biomarker measurements were available for a patient between T2DM incidence and earlier of subsequent CVD diagnosis or final visit/interview, the average of all available measurement values was used. Cochran's Q-test was used to determine whether the adjusted HR estimates from the cohorts were significantly different. IRs and HRs were estimated treating the death as a competing risk. We compared covariate missingness probabilities by CVD outcome status in both cohorts. We presented primary results from complete data analysis as EHR data typically violates the missing-at-random assumption required for most imputation techniques.^{5,6}

We used Harrell *c*-statistics to assess the discriminatory ability of each risk score within each applicable cohort and calibration plots to assess the performance of each risk score to predict AR. For each risk function, we divided the cohorts into deciles of predicted 5-year risk of the corresponding CVD outcome and calculated mean predicted risk within each decile. We calculated the observed Kaplan–Meier estimate within each decile and plotted the observed vs. predicted risk estimates.³⁹

Analyses were conducted using R version 1.2.5019.

RESULTS

The NYULH-EHR analysis flowchart is shown in [Figure 1A](#). Overall, 105 793 NYULH ambulatory patients were eligible and 19 849 (19.4%) had a query-based T2DM diagnosis ([Supplementary Table S1](#) and [Figure S1](#)). We performed sensitivity analyses by modifying key components of this algorithm, including the number of diagnostic codes and T2DM medication types ([Supplementary Table S1](#)) and found that T2DM prevalence varied between 18.2 and 20.2% in NYULH-EHR ambulatory patients age ≥ 50 . The NYULH-EHR

T2DM sample included 7432 unique patients with 130 387 encounters (median 12 per patient). Of these, 1773 (23.9%) developed subsequent CVD, with an 8.3-year maximum follow-up since T2DM diagnosis (median 2.6 years, interquartile range [IQR] 1.9–3.9 years). Among NYULH-EHR patients, 1050 (14.1%) developed hard CVD outcomes.

The HRS analysis flowchart is displayed in [Figure 1B](#). Among 37 610 HRS respondents with self-reported T2DM and CVD responses, 33 016 respondents denied T2DM at baseline. 3032 subsequently developed T2DM with no CVD at T2DM onset, and 667 (22.0%) developed subsequent CVD, with a 20-year maximum follow-up (median 8 years, IQR 4–12 years). There were 13 415 interviews for the 3032 respondents (median 4 per person).

[Table 1](#) presents characteristics of the NYULH-EHR and HRS subjects overall and by subsequent CVD status. Compared with HRS participants, NYULH-EHR T2DM patients were slightly younger, had a higher proportion of women and minority groups, and a lower proportion of smokers. Average SBP and HDL were similar between the two cohorts, while BMI, hypertension medication usage and HbA1c were higher within NYULH-EHR. In both cohorts, patients who developed CVD were more likely to be non-Hispanic Caucasians, more likely to smoke and had lower HDL levels. Within HRS, patients with subsequent CVD had higher BMI, SBP, and HbA1c, as well as more frequent antihypertensive medication use, but these differences were not significant within NYULH-EHR. Increased age was significantly associated with subsequent CVD in NYULH-EHR, but not within HRS. Sex was not associated with development of CVD in either cohort. Worse renal function was seen among those with CVD, demonstrated by lower albumin, lower eGFR, and higher creatinine in NYULH-EHR and higher CysC in HRS. Higher CRP was associated with CVD in HRS. In NYULH-EHR, CVD cases had more frequent encounters than controls, while sampling frequencies were fixed in HRS with bi-annual interviews.

In NYULH-EHR, 1% of samples had missing vitals, 8% missing HbA1c, 5–32% missing renal function biomarkers, and 13% missing lipid biomarkers. In HRS, ~35% of T2DM patients had missing biomarkers, consistent with the reported missing biomarker proportions in HRS as they were not sampled in the biomarker surveys.^{37,38} ([Supplementary Table S3](#)). In NYULH-EHR, CVD cases were more likely to have missing data than controls. In contrast, CVD controls had more missing data than cases in HRS.

Higher CVD IR estimates in NYULH-EHR

The estimated subsequent total CVD IR was 37.5 (95% confidence interval [CI] 34.8–40.5) and 90.6 (86.5–95.0) per 1000 person-years since T2DM onset in HRS and NYULH-EHR, respectively. We performed sensitivity analyses by modifying the required recurrent encounters with CVD ICD-10s and found that the subsequent CVD incidence ranged between 126.8 (requiring one encounter), 120.1 (two encounters) to 87.6 (four encounters) per 1,000 person-years, robustly higher than that estimated from HRS. The estimated subsequent hard CVD IR was 32.9 (95% CI 31.0–35.0) per 1000 person-years since T2DM onset in the NYULH-EHR.

Comparison of HR estimates for demographic covariates and biomarkers

Time-to-CVD from T2DM onset was analyzed using multivariable Cox proportional hazards models ([Table 2](#)). In both cohorts, increasing age, Caucasian race, smoking, elevated BMI, and lower

Table 1. Demographic characteristics of the analyzed NYULH-EHR T2DM patients and the HRS T2DM participants

CVD status	NYULH-EHR T2DM patients				HRS T2DM respondents			
	Overall (N = 7432)	No CVD (n = 5659)	CVD (n = 1773)	P-value	Overall (N = 3032)	No CVD (n = 2355)	CVD (n = 677)	P-value
Age at T2DM onset (years)	63.54 (10.25)	62.52 (9.93)	66.79 (10.58)	<0.001	65.61 (8.69)	65.61 (8.77)	65.89 (8.38)	0.99
Sex: female (%)	4495 (60.5)	3443 (60.8)	1052 (59.3)	0.269	1684 (55.5)	1326 (56.3)	358 (52.9)	0.12
Race (%)				<0.001				<0.001
Caucasian (White)	3598 (48.4)	2489 (44.0)	1109 (62.5)		2142 (70.6)	1606 (68.2)	536 (79.2)	
African American (Black)	1427 (19.2)	1208 (21.3)	219 (12.4)		646 (21.3)	529 (22.5)	117 (17.3)	
Other	1903 (25.6)	1526 (27.0)	377 (21.3)		241 (7.9)	217 (9.2)	24 (3.5)	
Unknown	504 (6.8)	436 (7.7)	68 (3.8)		3 (0.1)	3 (0.1)	0 (0.0)	
Ethnicity (%)				<0.001				<0.001
Hispanic or Latino	1498 (20.2)	1217 (21.5)	281 (15.8)		495 (16.3)	415 (17.6)	80 (11.8)	
Not Hispanic or Latino	5049 (67.9)	3699 (65.4)	1350 (76.1)		2537 (83.7)	1940 (82.4)	597 (88.2)	
Unknown	885 (11.9)	743 (13.1)	142 (8.0)		–	–	–	
Smoking (%)				0.024				<0.001
No	6965 (93.7)	5324 (94.1)	1641 (92.6)		1734 (57.2)	1265 (53.7)	469 (69.3)	
Yes	467 (6.3)	235 (5.9)	132 (7.4)		456 (15.0)	336 (14.3)	120 (17.7)	
Unknown	–	–	–		842 (27.8)	754 (32.0)	88 (13.0)	
Body mass index (BMI) (kg/m ²)	31.36 (6.31)	31.29 (6.27)	31.60 (6.44)	0.07	30.05 (6.06)	29.92 (6.00)	30.47 (6.25)	0.04
Systolic blood pressure (SBP) (mmHg)	130.47 (11.67)	130.57 (11.38)	130.15 (12.58)	0.19	132.46 (18.60)	131.74 (18.33)	134.67 (19.25)	0.002
High-density lipoprotein (HDL) (mg/dL)	51.99 (14.79)	52.43 (14.81)	50.36 (14.61)	<0.001	51.00 (13.84)	51.49 (13.93)	49.53 (13.50)	0.008
Hemoglobin A1c (HbA1c) (%)	7.18 (1.39)	7.16 (1.37)	7.25 (1.48)	0.018	6.71 (1.13)	6.68 (1.14)	6.81 (1.11)	0.03
Albumin (g/dL)	4.23 (0.33)	4.25 (0.33)	4.15 (0.35)	<0.001	–	–	–	
Creatinine (mg/dL)	0.94 (0.37)	0.92 (0.34)	1.02 (0.46)	<0.001	–	–	–	
eGFR (mL/min/1.73m ²)	74.00 (21.33)	76.28 (20.58)	64.77 (21.86)	<0.001	–	–	–	
Cystatin C (mg/dL)	–	–	–		1.21 (0.54)	1.16 (0.49)	1.37 (0.64)	<0.001
C-reactive protein (CRP) (mg/L)	–	–	–		4.52 (6.68)	4.15 (6.34)	5.62 (7.54)	<0.001
Anti-hypertensive medication use (%)				0.539				<0.001
No	1951 (26.3)	1496 (26.4)	455 (25.7)		76 (2.5)	75 (3.2)	1 (0.1)	
Yes	5481 (73.7)	4163 (73.6)	1318 (74.3)		1938 (63.9)	1488 (63.2)	450 (66.5)	
Unknown	–	–	–		1018 (33.6)	792 (33.6)	226 (33.4)	

(continued)

Table 1. continued

CVD status	NYULH-EHR T2DM patients				HRS T2DM respondents			
	Overall (N = 7432)	No CVD (n = 5659)	CVD (n = 1773)	P-value	Overall (N = 3032)	No CVD (n = 2355)	CVD (n = 677)	P-value
Diabetes medication use (%)				<0.001				<0.001
No	1957 (26.3)	1391 (24.6)	566 (31.9)		403 (13.3)	379 (16.1)	24 (3.5)	
Yes	5475 (73.7)	4268 (75.4)	1207 (68.1)		2381 (78.5)	1802 (76.5)	579 (85.5)	
Unknown	–	–	–		248 (8.2)	174 (7.4)	74 (10.9)	
Atherosclerotic CVD medication use (%)				0.312				–
No	2748 (37.0)	2074 (36.6)	674 (38.0)		–	–	–	
Yes	4684 (63.0)	3585 (63.4)	1099 (62.0)		–	–	–	
Encounters/year (median [IQR])	6.3 (3.9, 10.4)	5.6 (3.6, 9.0)	9.6 (5.9, 18.1)	<0.001	–	–	–	

Demographics and risk factors were summarized using means (standard deviations) for continuous measures and frequencies (percentages) for categorical measures. Comparisons were performed using the two-sample *t*-test and the chi-squared test, respectively. Biomarkers are summarized as the mean of all available biomarkers measurements following T2DM onset. Medications are summarized as prescriptions or self-reported medication usage at any encounter/visit following T2DM onset.

Table 2. Demographic covariates adjusted Cox models of time-to-CVD

Variables	NYULH-EHR		HRS		Test for heterogeneity P-value
	Hazard ratio [95% CI]	P-value	Hazard ratio [95% CI]	P-value	
Model 1					
Age at type 2 diabetes mellitus (T2DM) onset	1.03 [1.03, 1.04]	<0.001	1.03 [1.02, 1.04]	<0.001	0.34
Sex (reference = Male)					
Female	0.79 [0.71, 0.87]	<0.001	0.90 [0.77, 1.04]	0.15	0.18
Race (reference = Caucasian)					
African American (Black)	0.55 [0.47, 0.64]	<0.001	0.73 [0.60, 0.90]	0.003	0.03
Other	0.77 [0.67, 0.89]	<0.001	0.55 [0.36, 0.83]	0.005	0.12
Ethnicity (reference = Not Hispanic or Latino)					
Hispanic or Latino	0.60 [0.52, 0.70]	<0.001	0.72 [0.57, 0.92]	0.009	0.21
Smoking (reference = Non-smokers)					
Smokers	1.49 [1.23, 1.79]	<0.001	1.29 [1.05, 1.58]	0.015	0.32
Models 2–11					
Body mass index (BMI)	1.19 [1.13, 1.25]	<0.001	1.15 [1.06, 1.24]	0.001	0.46
Systolic blood pressure (SBP)	1.03 [0.97, 1.08]	0.359	1.06 [0.97, 1.16]	0.19	0.52
Total cholesterol (TC)	1.09 [1.03, 1.16]	0.003	0.95 [0.86, 1.05]	0.301	0.02
High-density lipoprotein (HDL)	0.84 [0.79, 0.90]	<0.001	0.89 [0.80, 0.99]	0.04	0.33
Hemoglobin A1c (HbA1c)	1.26 [1.19, 1.33]	<0.001	1.08 [0.98, 1.18]	0.11	0.01
Albumin	0.89 [0.84, 0.94]	<0.001	–	–	–
Creatinine	1.09 [1.05, 1.14]	<0.001	–	–	–
eGFR	0.77 [0.71, 0.84]	<0.001	–	–	–
Cystatin C	–	–	1.16 [1.09, 1.23]	<0.001	–
C-reactive protein	–	–	1.15 [1.09, 1.22]	<0.001	–

Model 1 is a Cox model with demographic covariates including age at T2DM diagnosis, sex, race, ethnicity, smoking status and, for NYULH-EHR, number of encounters. Models 2 to 11 are Cox models with covariates in Model 1 and one biomarker per model. Each biomarker was modeled separately, adjusting for the demographic covariates in Model 1 to accommodate not all biomarkers were available for the two datasets. HRs are per 1 standard deviation increase in the continuous covariates. The number of encounters was significant in more models. When multiple measurements were available for a patient longitudinally for a biomarker, the average of all available measurements values was used. HRs were estimated treating death as a competing risk. Cochran's Q-test was used to determine whether the adjusted HR estimates from the datasets were significantly different for each covariate.

HDL were associated with subsequent CVD, while increased SBP was not significantly associated with CVD development in either cohort. Worse renal function, indicated by higher creatinine, lower albumin, and lower eGFR in NYULH-EHR, and higher CysC in HRS,

was associated with worsened CVD risk. Elevated CRP levels also were associated with CVD in HRS. In both cohorts, AA race and Hispanic/Latino ethnicity were associated with decreased CVD risk. Females demonstrated a decreased CVD risk in both cohorts, but

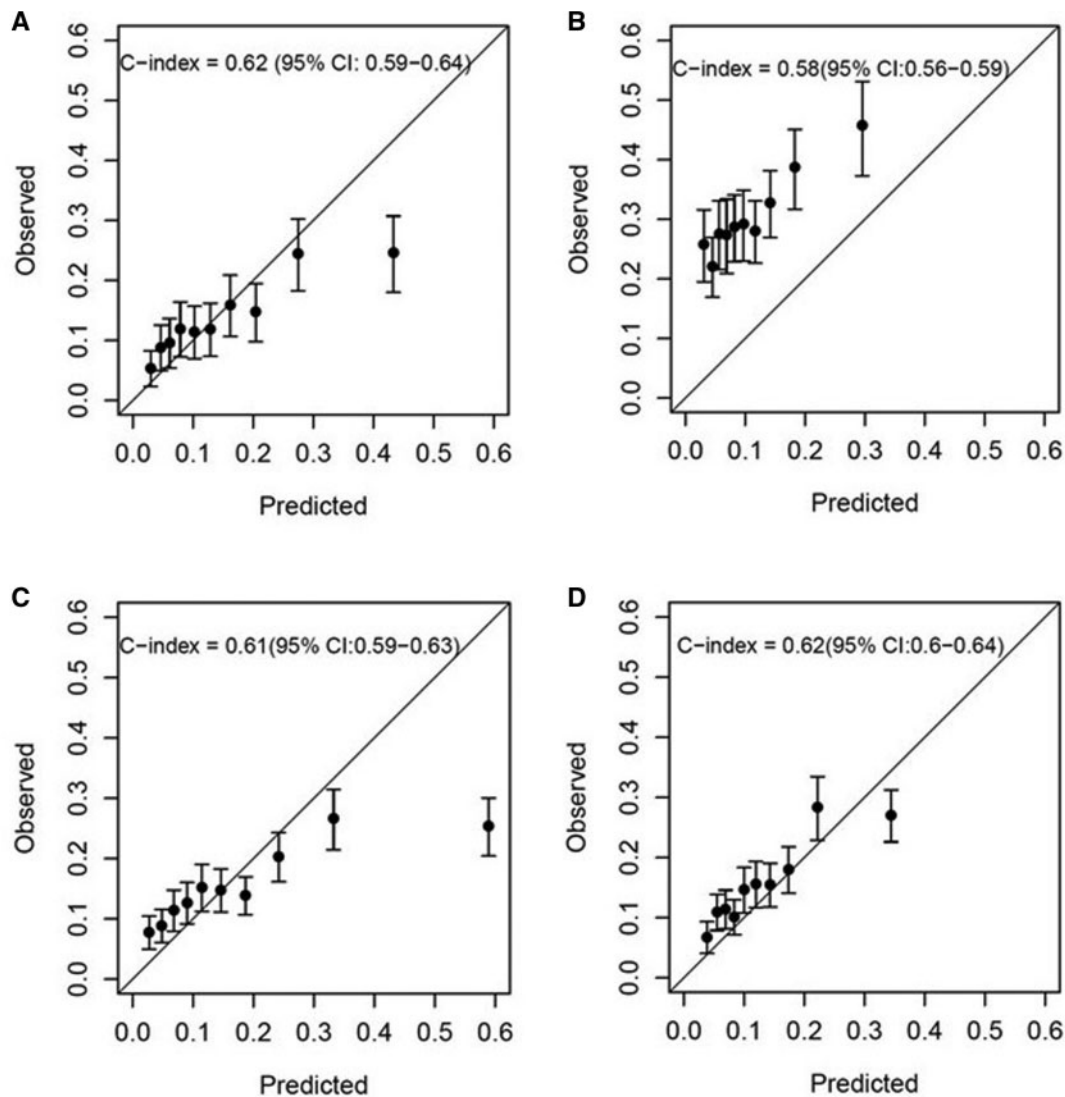


Figure 2. Calibration plots comparing predicted and observed 5-year risks. (A) HRS cohort using Framingham risk score to predict total CVD outcome; (B) NYULH-EHR cohort using Framingham risk score to predict total CVD outcome; (C) NYULH-EHR cohort using ACC/AHA pooled cohort equations to predict hard CVD outcome; (D) NYULH-EHR cohort using Swedish NDR to predict hard CVD outcome. Risk factors included in the FRS global CVD function and the ACC/AHA function are age, TC, HDL, SBP, treatment for hypertension, smoking, and T2DM status (all yes). Sex-specific risk equations were applied to males and females separately. For the ACC/AHA risk score, African American (AA) coefficients were used for AAs and white coefficients were used for all other patients. We replaced 10-year baseline survival estimates with 5-year estimates by assuming exponential survival distributions to align with the available follow-up of the present cohorts. Risk factors included in the Swedish NDR risk prediction functions include onset age of T2DM, T2DM duration, sex, BMI, smoking, HbA1c, SBP, and anti-hypertensive and lipid-reducing drugs. We divided the cohorts into deciles of the predicted risk and calculated the mean predicted risk value within each decile. We calculated the observed risk from the Kaplan–Meier estimate within each decile and plotted the observed vs. predicted risk functions.

this was only significant within NYULH-EHR. Lower TC and HbA1c were associated with decreased CVD risk only in NYULH-EHR. HRs of most covariates showed no significant differences between the two datasets, except AA race, TC, and HbA1c. As a sensitivity analysis, we analyzed NYULH-EHR models without adjustment for the number of encounters ([Supplementary Table S4](#)) and these results were robust.

Under-estimation of total CVD risk in NYULH-EHR

Calibration plots comparing predicted and observed CVD risks are shown in [Figure 2](#). The FRS had comparable discrimination ability for total CVD risk in HRS (c-index 0.62, [Figure 2A](#)) and NYULH-EHR (c-index 0.58, [Figure 2B](#)). In NYULH-EHR, the predicted total

CVD risks underestimated the observed CVD risks in all deciles due to the higher observed subsequent total CVD incidence. For hard CVD outcomes in NYULH-EHR, the ACC/AHA risk score ([Figure 2C](#)) showed discrimination with c-index 0.61 and the Swedish NDR risk score ([Figure 2D](#)) showed similar discrimination with c-index 0.62. However, there was no systematic under-estimation of ACC/AHA or Swedish NDR risk scores compared to observed hard CVD risks in NYULH-EHR.

DISCUSSION

The fundamentally different sampling mechanisms in EHRs compared to traditional cohort studies were often ignored, although it

has been acknowledged that this may lead to biased associations.⁴ In this manuscript, we performed the first systematic investigation on how different sampling mechanisms impact their inferences of incidence rates, relative risks, and absolute risks of the same outcome in the same target population. For this purpose, we compared the estimates of subsequent CVD incidence and HRs of demographic/clinic risk factors with CVD outcomes obtained from the T2DM patients in NYULH-EHR hospital-based ambulatory care and from the T2DM patients in HRS.

We found that the estimated incidence of subsequent total CVD was substantially higher for NYULH-EHR T2DM patients than HRS T2DM participants, even with varying numbers of required recurrent diagnostic codes in the EHR phenotyping algorithm.^{5–7} Total CVD risk predictions, established from traditional cohort studies, consequently underestimated the observed total CVD risks in NYULH-EHR, yet not in HRS. This consistent with the literature that hospital-based EHR samples are not random samples from the population, even for disease-specific populations, as they tend to be sicker and with higher healthcare utilization.^{5,6,40} In contrast, the HRS CVD identification relies on adjudicated self-reported heart disease or stroke, which may be biased by inaccurate recall, underdiagnosis, or inclusion of healthier patients.⁸

Despite these differences, traditional demographic and biomarker CVD risk factors showed comparable risk estimates in NYULH-EHR and HRS. The HR estimates and c-indices were consistent with what was reported from the literature,^{20,41–43} providing credence of the utility of a well-curated single-institution EHR sample as a feasible cohort to investigate risk factors and refine risk prediction models. Although there was no direct comparison of hard CVD outcomes because no details in CVD outcomes were collected in HRS, we did not observe systemic overestimation for hard CVD outcomes in NYULH-EHR versus ACA/AHA and Swedish NDR risk predictions. This indicates hard CVD outcome is more robust and consistent between EHRs and cohort studies, which may be a result of hard CVD outcomes being diagnosed more consistently between providers than softer CVD outcomes.

Several additional differences between the two datasets were observed. First, the frequency of encounters was significantly higher in patients with CVD in NYULH-EHR, whereas this was fixed by design in HRS.^{5,6} It therefore is important to adjust for this as a potential confounding effect and to compare its impact with the unadjusted model in EHR association analysis. These indicators are treated as confounding variables and are not meant to be causal variables or interpreted as such.⁴⁴ Second, missingness is higher in CVD patients in NYULH-EHR while lower in CVD patients in HRS, both of which are associated with outcome status. The observed differences in missing probability may be caused by the differences in follow-up time. In NYULH-EHR, patients with CVD had a shorter time to event than those with no CVD and therefore had fewer lab results available. However, in HRS, participants with longer follow-up time were more likely to observe CVD outcomes, and more likely to be included in the biomarker substudy. Neither sample satisfies missing-at-random assumptions, therefore specific methods to handle missing data are warranted.

Possibly unintuitive findings, yet consistently supported by both datasets, were as following. First, the FRS had limited discrimination c-indices for total CVD events in both two datasets, and the ACC/AHA PCE and Swedish NDR risk scores had limited c-indices for hard CVD events in NYULH-EHR. Our population is an older, exclusively diabetic cohort, which differs from the general population cohorts used to develop the FRS, ACC/AHA, and Swedish

NDR risk scores. Therefore, the estimated c-indices are presented to compare performance in the two data sources, not to conclude the invalidity of these risk scores. However, the moderate discriminations and under-estimation of total CVD risk for older ambulatory T2DM patients warrants further investigation as US and international guidelines for primary prevention of CVD have all recommended the use of global CVD absolute risk assessment for decision-making about the intensity of lifestyle and pharmacological preventive interventions.^{45–49}

Both cohorts showed lower subsequent CVD incidence for Hispanic and non-Caucasian T2DM patients despite their higher T2DM prevalence, consistent with previous findings termed the “Hispanic Paradox”.^{50,51} The observed difference was not explained by heterogeneous CVD prevalence before or at the time of T2DM diagnosis, nor by the incidence of death as a competing outcome. One plausible explanation is the underdiagnosis/reporting of CVD outcomes for minority T2DM subjects, especially in a hospital-based EHR sampling scheme.⁵ In NYULH-EHR, minority T2DM patients had shorter follow-up (Hispanic median 1.3, AA 2.4, Caucasian 3.5 years since T2DM diagnosis) and fewer encounters. AA and Hispanics showed significant protective HRs in the encounter-adjusted results. In HRS, follow-up is comparable by design with slightly fewer interviews among minority patients (average 3.7, 4.1, and 4.5 bi-annual interviews since T2DM onset for Hispanics, AA, and Caucasians, respectively). These findings support a need for greater understanding of data reliability and potential racial/ethnicity confounding factors CVDs to appropriately model CVD prevalence, incidence, risk, and resilience, and improve targeting of intervention efforts for racial/ethnicity minorities.⁵⁰

Our study has several limitations. First, there are secular differences in T2DM and CVD management between the NYULH-EHR and HRS data collection years. We observed higher HbA1c in NYULH-EHR compared to HRS (7.18% vs. 6.65%) as more recent guidelines from the American College of Physicians have recommended targeting HbA1c levels between 7% and 8%, whereas the previous target had been 6.5%.⁵² Second, the HRS time-to-CVD outcome is in 2-year units as it is obtained from bi-annual interviews, while the NYULH-EHR time-to-outcome has a more real-time resolution. However, the NYULH-EHR sample has a shorter follow-up period than the HRS sample, so our risk estimates may not be capturing later incident CVDs. Third, NYULH-EHR was from a single ambulatory healthcare system in the New York City metropolitan area. These findings do not apply to EHR data from universal-registration of national single-payer system such as in the United Kingdom or Scottish Care Information-Diabetes (SCI-Diabetes database), which present the overall national population under care. Finally, our assessment of CVD risk prediction functions was limited to risk scores utilizing measurements available in the NYULH-EHR and HRS cohorts, and as such may not be reflective of algorithms more commonly used in clinical practice such as the Multi-Ethnic Study of Atherosclerosis (MESA) risk score. Future studies at additional ambulatory care centers with longer follow-up and additional clinical feature collection certainly are warranted to confirm these findings.

To our knowledge, this is the most comprehensive study of incident CVD among older chronic T2DM patients in a hospital-based ambulatory care setting in direct comparison to a US population-based longitudinal cohort. Being able to compare NYULH-EHR to HRS and ultimately demonstrating that a single institution’s EHR-based cohort estimated relative risks reasonably well is a crucial first step towards utilizing EHRs for general population inference.

CONCLUSION

Subsequent total CVD IR was higher in the T2DM patients of a hospital-based ambulatory care EHR cohort than the T2DM respondents of the national cohort Health and Retirement Survey. Therefore, the FRS prediction for total CVD outcomes underestimated the observed CVD risks in the T2DM EHR samples. However, the HRs of demographic factors (including age, gender, race, ethnicity, and smoking status) and major clinical risk factors were generally concordant between the two cohorts.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONTRIBUTIONS

J.S. compiled and interpreted the HRS data; primary contributor to the written manuscript. S.C. compiled and interpreted the NYULH-EHR data. C.C., J.H., M.K., and J.Y. assisted in data interpretation and contributed significantly to writing and editing manuscript. J.D., L.T., and C.B. reviewed all data analyses and edited the final manuscript. J.Z. coordinated interaction between authors, designed the overall study, and performed all data analyses. All authors have read and approved the final manuscript, contributing edits where applicable.

CONFLICT OF INTEREST STATEMENT

None declared.

FUNDING

J.S., C.C., J.H., M.K., J.Y., J.D., C.B., and J.Z. are funded by National Institute on Aging (NIA) R01AG054467 and NIA R01AG065330.

REFERENCES

- Nichols GA, Desai J, Elston Lafata J, *et al.*; SUPREME-DM Study Group. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM project. *Prev Chronic Dis* 2012; 9: E110.
- Catalan-Ramos A, Verdu JM, Grau M, *et al.*; GPC-ICS Group. Population prevalence and control of cardiovascular risk factors: what electronic medical records tell us. *Aten Primaria* 2014; 46 (1): 15–24.
- Sidebottom AC, Johnson PJ, VanWormer JJ, Sillah A, Winden TJ, Boucher JL. Exploring electronic health records as a population health surveillance tool of cardiovascular disease risk factors. *Popul Health Manag* 2015; 18 (2): 79–85.
- Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017; 24 (1): 198–208.
- Goldstein BA, Bhavsar NA, Phelan M, Pencina MJ. Controlling for informed presence bias due to the number of health encounters in an electronic health record. *Am J Epidemiol* 2016; 184 (11): 847–55.
- Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA Annu Symp Proc* 2013; 2013: 1472–7.
- Bower JK, Patel S, Rudy JE, Felix AS. Addressing bias in electronic health record-based surveillance of cardiovascular disease risk: finding the signal through the noise. *Curr Epidemiol Rep* 2017; 4 (4): 346–52.
- Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J* 2003; 20 (1): 54–60.
- Al-Salameh A, Chanson P, Bucher S, Ringa V, Becquemont L. Cardiovascular disease in type 2 diabetes: a review of sex-related differences in predisposition and prevention. *Mayo Clin Proc* 2019; 94 (2): 287–308.
- Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc Diabetol* 2018; 17 (1): 83.
- Moreno G, Mangione CM. Management of cardiovascular disease risk factors in older adults with type 2 diabetes mellitus: 2002–2012 literature review. *J Am Geriatr Soc* 2013; 61 (11): 2027–37.
- Halter JB, Musi N, McFarland Horne F, *et al.* Diabetes and cardiovascular disease in older adults: current status and future directions. *Diabetes* 2014; 63 (8): 2578–89.
- Sazlina SG, Mastura I, Ahmad Z, *et al.* Control of glycemia and other cardiovascular disease risk factors in older adults with type 2 diabetes mellitus: data from the Adult Diabetes Control and Management. *Geriatr Gerontol Int* 2014; 14 (1): 130–7.
- IDF Diabetes Atlas*. Brussels, Belgium: International Diabetes Federation; 2015.
- Einarson TR, Acs A, Ludwig C, Panton UH. Economic burden of cardiovascular disease in type 2 diabetes: a systematic review. *Value Health* 2018; 21 (7): 881–90.
- Dinesh Shah A, Langenberg C, Rapsomaniki E, *et al.* Type 2 diabetes and incidence of a wide range of cardiovascular diseases: a cohort study in 1.9 million people. *Lancet* 2015; 385 (Suppl 1): S86.
- Shah AD, Langenberg C, Rapsomaniki E, *et al.* Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1.9 million people. *Lancet Diabetes Endocrinol* 2015; 3 (2): 105–13.
- Gregg EW, Li Y, Wang J, *et al.* Changes in diabetes-related complications in the United States, 1990–2010. *N Engl J Med* 2014; 370 (16): 1514–23.
- Goff DC, L-J D, Bennett G, *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association task force on practical guidelines. *Circulation* 2014; 129 (25 suppl 2): S49–S73.
- McClelland RL, J N, Budoff M, *et al.* Ten-year coronary heart disease risk prediction using coronary artery calcium and traditional risk factors: derivation in the multi-ethnic study of atherosclerosis with validation in the Heinz Nixdorf Recall Study and the Dallas Heart Study. *J Am Coll Cardiol* 2015; 66 (15): 1643–53.
- Wilson PWF, D’Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; 97 (18): 1837–47.
- Koller MT, L M, Wolbers M, *et al.* Development and validation of a coronary risk prediction model for older U.S. and European persons in the Cardiovascular Health Study and the Rotterdam Study. *Ann Intern Med* 2012; 157 (6): 389–97.
- D’Agostino RB, Vasan RS, Pencina MJ, *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; 117 (6): 743–53.
- Hippisley-Cox J, Coupland C, Vinogradova Y, *et al.* Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008; 336 (7659): 1475–82.
- Kengne AP, Patel A, Marre M, *et al.*; ADVANCE Collaborative Group. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. *Eur J Cardiovasc Prev Rehabil* 2011; 18 (3): 393–8.
- Zethelius B, Eliasson B, Eeg-Olofsson K, Svensson A-M, Gudbjörnsdottir S, Cederholm J. NDR. A new model for 5-year risk of cardiovascular disease in type 2 diabetes, from the Swedish National Diabetes Register (NDR). *Diabetes Res Clin Pract* 2011; 93 (2): 276–84.
- Read SH, vD M, Colhoun HM, Halbesma N, *et al.*; Scottish Diabetes Research Network Epidemiology Group. Performance of cardiovascular disease risk scores in people diagnosed with type 2 diabetes: external validation using data from the National Scottish Diabetes Register. *Diabetes Care* 2018; 41 (9): 2010–8. ;

28. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the Health and Retirement Study (HRS). *Int J Epidemiol* 2014; 43 (2): 576–85.
29. Health and Retirement Study (HRS) [article online], Available from <https://www.nia.nih.gov/research/resource/health-and-retirement-study-hrs>. Accessed 08/11/2020.
30. Kelley AS, Langa KM, Smith AK, *et al.* Leveraging the health and retirement study to advance palliative care research. *J Palliat Med* 2014; 17 (5): 506–11.
31. Hong I, Reistetter TA, Díaz-Venegas C, Michaels-Obregon A, Wong R. Cross-national health comparisons using the Rasch model: findings from the 2012 US Health and Retirement Study and the 2012 Mexican Health and Aging Study. *Qual Life Res* 2018; 27 (9): 2431–41.
32. Juster FT, Suzman R. An overview of the Health and Retirement Study. *J Hum Resour* 1995; 30: S7–S56.
33. Fisher GG, Ryan LH. Overview of the Health and Retirement Study and introduction to the special issue. *Work, Aging Retire* 2018; 4 (1): 1–9.
34. Cigolle CT, Nagel CL, Blaum CS, Liang J, Quiñones AR. Inconsistency in the self-report of chronic diseases in panel surveys: developing an adjudication method for the Health and Retirement Study. *J Gerontol B Psychol Sci Soc Sci* 2018; 73 (5): 901–12.
35. Elley CR, Robinson E, Kenealy T, Bramley D, Drury PL. Derivation and validation of a new cardiovascular risk score for people with type 2 diabetes: the New Zealand Diabetes Cohort Study. *Diabetes Care* 2010; 33 (6): 1347–52.
36. Pylypchuk R, Wells S, Kerr A, *et al.* Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. *Lancet* 2018; 391 (10133): 1897–907.
37. Documentation of biomarkers in the 2006 and 2008 Health and Retirement Study [article online], 2013. <http://hrsonline.isr.umich.edu/modules/meta/bio2008/desc/Biomarker2006and2008.pdf> Accessed December 16, 2019.
38. Documentation of biomarkers in the 2010 and 2012 Health and Retirement Study [article online], 2015. <http://hrsonline.isr.umich.edu/modules/meta/bio2012/desc/Biomarker2010and2012.pdf> Accessed December 16, 2019.
39. Triant VA, Perez J, Regan S, *et al.* Cardiovascular risk prediction functions underestimate risk in HIV infection. *Circulation* 2018; 137 (21): 2203–14.
40. Ruano-Ravina A, Pérez-Ríos M, Miguel Barros-Dios J. Population-based versus hospital-based controls: are they comparable? *Gaceta Sanitaria* 2008; 22 (6): 609–13.
41. Tsao CW, Vasan RS. Cohort profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol* 2015; 44 (6): 1800–13.
42. Cederholm J, Eeg-Olofsson K, Eliasson B, Zethelius B, Nilsson PM, Gudbjornsdottir S; on behalf of the Swedish National Diabetes Register. Risk prediction of cardiovascular disease in type 2 diabetes: a risk equation from the Swedish National Diabetes Register. *Diabetes Care* 2008; 31 (10): 2038–43.
43. Hashemi Madani N, Ismail-Beigi F, Khamseh ME, Malek M, Ebrahimi Valojerdi A. Predictive and explanatory factors of cardiovascular disease in people with adequately controlled type 2 diabetes. *Eur J Prev Cardiol* 2017; 24 (11): 1181–9.
44. Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton: Chapman & Hall/CRC, 2008.
45. Andrus B, Lacaille D. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. *J Am Coll Cardiol* 2014; 63 (25): 2886.
46. National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002; 106: 3143–421.
47. Perk J, De Backer G, Gohlke H, *et al.*; European Association for Cardiovascular Prevention & Rehabilitation, ESC Committee for Practice Guidelines. European guidelines on cardiovascular disease prevention in clinical practice (version 2012), The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). *Eur Heart J* 2012; 33 (13): 1635–701.
48. Anderson TJ, Gregoire J, Hegele RA, *et al.* 2012 update of the Canadian Cardiovascular Society guidelines for the diagnosis and treatment of dyslipidemia for the prevention of cardiovascular disease in the adult. *Can J Cardiol* 2013; 29 (2): 151–67.
49. Genest J, McPherson R, Frohlich J, Anderson T, *et al.* 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult - 2009 recommendations. *Can J Cardiol* 2009; 25 (10): 567–79.
50. Balfour PC Jr, Ruiz JM, Talavera GA, Allison MA, Rodriguez CJ. Cardiovascular disease in Hispanics/Latinos in the United States. *J Lat Psychol* 2016; 4 (2): 98–113.
51. Rodriguez CJ, Allison M, Daviglius ML, *et al.*; American Heart Association Council on Epidemiology and Prevention, American Heart Association Council on Clinical Cardiology, American Heart Association Council on Cardiovascular and Stroke Nursing. Status of cardiovascular disease and stroke in Hispanics/Latinos in the United States: a science advisory from the American Heart Association. *Circulation* 2014; 130 (7): 593–625.
52. Qaseem A, Wilt TJ, Kansagara D, Horwitch C, Barry MJ, Forciea MA; Clinical Guidelines Committee of the American College of Physicians. Hemoglobin A1c targets for improving glycemic control with pharmacologic therapy for nonpregnant adults with type 2 diabetes mellitus: a guidance statement update from the American College of Physicians. *Ann Intern Med* 2018; 168 (8): 569–76.