Development and Validation of a Risk Stratification Model Using Disease Severity Hierarchy for Mortality or Major Cardiovascular Event

Che Ngufor, PhD; Pedro J. Caraballo, MD; Thomas J. O’ Byrne, MAS; David Chen, PhD; Nilay D. Shah, PhD; Lisiane Pruinelli, PhD, RN; Michael Steinbach, PhD; Gyorgy Simon, PhD

Abstract

IMPORTANCE Clinical domain knowledge about diseases and their comorbidities, severity, treatment pathways, and outcomes can facilitate diagnosis, enhance preventive strategies, and help create smart evidence-based practice guidelines.

OBJECTIVE To introduce a new representation of patient data called disease severity hierarchy that leverages domain knowledge in a nested fashion to create subpopulations that share increasing amounts of clinical details suitable for risk prediction.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study included 51,969 patients aged 45 to 85 years, with 10,674 patients who received primary care at the Mayo Clinic between January 2004 and December 2015 in the training cohort and 41,295 patients who received primary care at Fairview Health Services from January 2010 to December 2017 in the validation cohort. Data were analyzed from May 2018 to December 2019.

MAIN OUTCOMES AND MEASURES Several binary classification measures, including the area under the receiver operating characteristic curve (AUC), Gini score, sensitivity, and positive predictive value, were used to evaluate models predicting all-cause mortality and major cardiovascular events at ages 60, 65, 75, and 80 years.

RESULTS The mean (SD) age and proportions of women and white individuals were 59.4 (10.8) years, 6324 (59.3%) and 9804 (91.9%), respectively, in the training cohort and 57.4 (7.9) years, 21,975 (53.1%), and 37,653 (91.2%), respectively, in the validation cohort. During follow-up, 945 patients (8.9%) in the training cohort died, while 787 (7.4%) had major cardiovascular events. Models using the new representation achieved AUCs for predicting death in the training cohort at ages 60, 65, 75, and 80 years of 0.96 (95% CI, 0.94-0.97), 0.96 (95% CI, 0.95-0.98), 0.97 (95% CI, 0.96-0.98), and 0.98 (95% CI, 0.97-0.99), respectively, while standard methods achieved modest AUCs of 0.67 (95% CI, 0.55-0.80), 0.66 (95% CI, 0.56-0.79), 0.64 (95% CI, 0.57-0.71), and 0.63 (95% CI, 0.54-0.70), respectively.

CONCLUSIONS AND RELEVANCE In this study, the proposed patient data representation accurately predicted the age at which a patient was at risk of dying or developing major cardiovascular events substantially better than standard methods. The representation uses known relationships contained in electronic health records to capture disease severity in a natural and clinically meaningful way. Furthermore, it is expressive and interpretable. This novel patient representation can help to support critical decision-making, develop smart guidelines, and enhance health care and disease management by helping to identify patients with high risk.
Introduction

More than one-quarter of US adults have multiple (ie, \( \geq 2 \)) chronic conditions (MCCs), with a prevalence that increases with age.\(^1,3^\) Patients with MCCs use more health care resources, incur higher costs of care, and face greater mortality, with worse quality of life.\(^3,5^\) However, identifying, treating, and improving health outcomes of patients with MCCs have been hindered by the marked complexity and heterogeneity of this population. Current comparative effectiveness research and evidence-based practice guidelines (EBPGs) often focus on 1 disease at a time or on highly prevalent co-occurring conditions, with risk stratification based on simple comorbidity counts.\(^6^\) However, patients with MCCs represent complex cases and could benefit from smarter EBPGs that leverage data-driven machine learning and data about disease interactions, disease severity, and treatment pathways.

Predictive analytics based on electronic health records (EHRs) can provide crucial information for making better clinical decisions about patients with MCCs. However, it also raises important questions about the appropriate representation of MCCs for predicting health outcomes. Current methods tend to focus on traditional comorbidity representations, such as simple summaries of individual conditions (eg, counts, sums, and most recent values) or aggregate measures of comorbidities.\(^6^\) Unfortunately, these representations do not reflect an individual’s comorbid disease history and severity and do not account for interactions with different treatment pathways. Therefore, this leaves a large gap in our knowledge regarding how to optimally manage complex MCCs.\(^7^\)

To address this critical knowledge gap, we introduce a novel representation of patient data, disease severity hierarchy (DSH), which explores diseases and their known treatment pathways in a nested fashion. As the DSH tree is traversed from the root to the leaves, subpopulations of patients with similar clinical characteristics, such as disease severity, progression, and survival, are created, thereby providing discriminative features suitable for developing risk stratification models. To quantitatively represent the information embedded in DSH, we further introduce a risk scoring system, which encodes disease severity and its intensification as numeric values along any path of the DSH tree. We then test the effectiveness of DSH in the case of type 2 diabetes and its comorbidities (ie, hypertension, obesity, and hyperlipidemia) in predicting all-cause mortality (ACM) and major cardiovascular events (MCEs).

Methods

This study was reviewed and approved by the Mayo Clinic institutional review board as a minimal risk study, and informed consent was not required. The study conforms to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.\(^8^\)

Study Source and Population

We retrospectively analyzed the medical records of 10,674 adults, aged 45 to 85 years, included in the Rochester Epidemiology Project (REP) database who received primary care at the Mayo Clinic from January 1, 2004, to December 31, 2015, to develop the models.\(^9^\) The models were externally validated using data for 41,295 patients, aged 45 to 85 years, who had a primary care visit between January 1, 2010, and December 31, 2017, at Fairview Health Services (FHS).\(^10^\) Patients in the REP data entered the study at their age on January 1, 2004 (index date), and were still alive on December 31, 2010. These patients were followed up until December 31, 2015. Correspondingly, patients in the FHS data were included in the analysis if they were aged at least 45 years on January 1, 2010 (index date), and were alive on December 31, 2015. The FHS patients were followed up until December 31, 2017. We defined the period between the index date and start of follow-up as the baseline time window (BTW). Figure 1A shows an overview of the study design.
Outcomes
The primary outcome was defined as the patient’s age-time of ACM and MCE in the follow-up period. Patients entered the study at age 45 years, were left truncated at age 51 years (REP) and age 50 years (FHS) when they were first classified as at risk, and exited the study at the event (ACM or MCE) or censored. The left truncation allowed us to put patients in the risk set only when they were actually under follow-up by the study. The use of the age-time scale provided an expressive and flexible way to control the effect of age, especially for older adults. It also provided a relatively meaningful basis on which to examine how risk varies over time.

Death records in the REP and FHS were obtained from the Minnesota Electronic Death Certificates and the National Death Index. We defined MCE as a composite of myocardial infarction, stroke, percutaneous transluminal coronary angioplasty, use of cardiac devices, coronary artery procedures, congestive heart failure, ischemic heart disease, coronary artery disease, cardiomyopathy, cardiac arrest, or angina occurring in the follow-up period. The list of International Classification of Diseases, Ninth Revision (ICD-9) and International Statistical Classification of...
Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes for these conditions can be found in the eAppendix in the Supplement. The REP and FHS have robust data quality standards, so missing values were low, and deaths and MCE records were complete in this study.

**Covariates**
Baseline demographic variables used for the analysis included sex and race/ethnicity. Comorbidities of type 2 diabetes, laboratory results, vital signs, and medication use were ascertained from the index date to the start of follow-up.

**Type 2 Diabetes and Its Comorbidities**
A patient was considered to have type 2 diabetes and its comorbidities if at least 1 of the following indications was present during BTW: (1) type 2 diabetes, clinical diagnosis (based on ICD-9 and ICD-10-CM codes), fasting blood glucose of at least 126 mg/dL (to convert to millimoles per liter, multiply by 0.0555), random blood glucose level of at least 200 mg/dL, or glycated hemoglobin (HbA1c) level of at least 6.5% (to convert to proportion of total hemoglobin, multiply by 0.01)\(^14\); (2) hypertension, clinical diagnosis or blood pressure at least 140/90 mm Hg; (3) hyperlipidemia, low-density lipoprotein cholesterol of at least 100 mg/dL (to convert to millimoles per liter, multiply by 0.0259); and (4) obesity, body mass index (calculated as weight in kilograms divided by height in meters squared) of at least 30.

**Summarizing Time-Varying Covariates**
The comorbidities, laboratory results, vital signs, and medications were captured throughout BTW; as such, a patient may have repeated measurements of the same variable. These potentially correlated measurements present a major challenge for conventional machine learning methods.\(^15,16\) Therefore, we aggregated the repeated observations for each patient to create cross-sectional data.

We created 2 comorbidity variables (the rolling comorbidity,\(^17\) ie, the condition persisted throughout the BTW, and the most frequent value), 2 medication variables (the number of times [frequency] the patient received any medication class and an indicator for the presence or absence of any class), and 3 variables for laboratory results and vital signs (frequency, median, and last observed value in BTW).

No observable association was found in the patterns of missing values (laboratory results and vital signs) in the data (eFigure 1 and eFigure 2 in the Supplement). Missing laboratory results and vital signs (with <35% missingness) were imputed using the random forest imputation method.\(^18\)

**DSH**
Diabetes and other heterogeneous diseases have multiple underlying molecular mechanisms that are reflected in varying sets of comorbidities.\(^19\) The heterogeneity can be largely viewed as manifestations of different disease mechanisms, which are present in different subpopulations and in different proportions. Unfortunately, the underlying disease mechanisms are not known a priori.

The proposed DSH is designed to address how heterogeneous diseases, such as diabetes, can be modeled without explicitly knowing the underlying disease mechanisms. It accomplishes this goal by transforming the multifaceted heterogeneous EHR data into a clinically meaningful patient state representation that uniquely captures and encodes known relationships about disease severity, treatment pathways, and outcomes. For each clinical encounter, we extracted diagnoses, laboratory results, vital signs, and interventions to construct the DSH representing the patient's health trajectory as a sequence of clinical findings.

**Clinical Findings and Actions**
We considered a patient's health trajectory a sequence of clinical findings and actions triggered by these findings. For instance, a new laboratory result (eg, a fasting blood glucose level of 130 mg/dL) and the corresponding clinical context describing the disease, its severity, symptoms, medications,
interventions, or outcomes observed during a clinical visit represent a clinical finding. After examining the finding, an action may be triggered (eg, introducing a new diabetic medication).

The sequence of clinical findings can be described at different granular levels: disease presence or absence, untreated disease, disease requiring therapy, aggressiveness of therapy (first-line, second-line, or last-line therapy), dose increase or decrease, and disease control status. Thus, DSH captures the patient’s health status, disease severity, treatment pathway, and outcomes, which can be represented as a binary tree. Figure 2 depicts a DSH representation for type 2 diabetes, and eFigure 3 in the Supplement depicts a DSH representation for obesity.

**Modeling Relationships in EHR**

Data from EHRs are rich with information on disease conditions, laboratory results, vital signs, medications, and treatment pathways. Unfortunately, most existing methods treat this information as independent and do not take advantage of preexisting domain knowledge about relationships. In contrast, DSH concisely captures this expert knowledge. Consider a patient taking metformin and sulfonylurea with an HbA₁c of 6%. At the sixth level of the DSH tree (Figure 2), the patient is controlled; at the fourth, is receiving second-line therapy; at the first, has diabetes.

**DSH Risk Score**

In traversing the DSH tree from the root to the leaves, we capture disease severity beneath each node; the right branch leads to intensification of the disease (eg, ≥2 drugs) and the left branch leads to a steady or controlled state (eg, first-line therapy). Motivated by this natural flow of information, we proposed a simple risk scoring system to quantitatively represent the information embedded in DSH. Specifically, we assigned risk scores to the nodes so that the score on the right branch doubles that of the node above, while the score on the left branch is the same as that of the node above. The risk score of the root node is either 1 (with disease) or 0 (disease free). We aggregated the time-varying DSH risk scores in the BTW into 3 DSH risk variables, as follows: sum, mean, and last observed value.

Figure 2. Disease Severity Hierarchy for Type 2 Diabetes

An example of a 6-level disease severity hierarchy tree structure and corresponding risk score allocations for type 2 diabetes. Disease severity intensifies on any right branch from the root to the leaves. A patient is considered under control if the laboratory result or vital sign associated with the condition is within its predefined reference range, eg, glycated hemoglobin less than 6.5% (to convert to proportion of total hemoglobin, multiply by 0.01). The disease severity hierarchy structure for hypertension and hyperlipidemia are similar to that for type 2 diabetes. An example of the disease severity hierarchy tree for obesity is presented in eFigure 3 in the Supplement.
Statistical Analysis

We report descriptive statistics of patient characteristics, with means and SDs for continuous variables and counts and percentages for categorical variables, and of the performance of the models. We developed age-specific survival models, in which we used the patient’s age instead of the traditional time-on-study as the scale for predicting the risk of ACM and MCE at ages 60, 65, 75, and 80 years based on DSH risk scores, while adjusting for sex and race/ethnicity. We trained the regularized Cox proportional hazard (CPH) regression model through 10-fold cross-validation.20 All data analyses were performed using R version 3.5.1 (R Project for Statistical Computing). No prespecified level of statistical significance was set.

Training and Validation

A regularized CPH algorithm requires the selection of 2 tuning parameters to avoid overfitting, as follows: (1) the least absolute shrinkage and selection operator (lasso) penalty and (2) the ridge penalty.21,22 We set up a grid for each combination and selected the best in 10-fold cross-validation. The final model was developed on the entire REP using the best parameters and validated externally on FHS (Figure 1B).

Performance Measures

We computed the accuracy, receiver operating characteristic curve, area under the receiver operating characteristic curve (AUC), cumulative gain curve and Gini score,23 sensitivity, specificity, and positive predictive value (PPV) at each age-time point using methods appropriate for censored data.24 The clinical utility of the models was evaluated using the net benefit computed across classification thresholds.25,26 The mean and 95% CIs of the performance scores over the cross-validations on REP data and point estimates on FHS data are reported. We show the association between sensitivity, PPV, and prediction density graphically with respect to the threshold.

Model Comparison

Because DSH captures laboratory results and medications associated with type 2 diabetes and its comorbidities, we conducted experiments to compare its predictive ability with that of the individual comorbidities, medications, laboratory results, and vital signs. Specifically, we compared the performance of models trained using the following group of predictors: DSH risk scores only (DSH-RS), comorbidities with medications (COM), and COM with laboratory results and vital signs (COM + LB/VS). We aimed to show that DSH-RS outperforms COM and had similar or better performance compared with COM + LB/VS.

Results

Baseline Characteristics

We implemented a 4-level DSH tree, in which the maximum allocated DSH risk score was 8. Table 1 presents descriptive statistics for the REP (10 674 participants) and FHS (41 295 participants) data. The REP data contained older patients and more women than the FHS data (mean [SD] age, 59.4 [10.8] years vs 57.4 [7.9] years; 6324 [59.3%] women vs 21 975 [53.2%] women). Both data sets comprised primarily white patients (REP, 9804 [91.9%]; FHS, 37 653 [91.2%]). During follow-up, 945 patients (8.9%) in the REP cohort died, while 787 (7.4%) had an MCE. Correspondingly, in the FHS cohort, 1857 (4.5%) died and 1857 (4.5%) had an MCE. The proportion of patients with missing laboratory results and vital signs in REP and FHS were 1.6% (171) and 0.8% (330), respectively (eFigure 1 and eFigure 2 in the Supplement).

Predicting ACM and MCE

Table 2 presents performance on REP and FHS. DSH-RS substantially outperformed COM. Specifically, models using DSH-RS achieved outstanding AUCs in predicting ACM at 60, 65, 75, and
Table 1. Study Population

| Variable                  | REP (n = 10 674) | FHS (n = 41 295) |
|---------------------------|------------------|------------------|
| **Demographic characteristic** |                  |                  |
| Age, mean (SD), yᵃ       | 59.4 (10.8)      | 57.4 (7.9)       |
| Women                     | 6324 (59.3)      | 21 975 (53.2)    |
| **Race/ethnicity**        |                  |                  |
| White                     | 9804 (91.9)      | 37 653 (91.2)    |
| Black                     | 179 (1.7)        | 1586 (3.8)       |
| Asian                     | 369 (3.5)        | 902 (2.2)        |
| American Indian           | 21 (0.2)         | 240 (0.6)        |
| Hawaiian                  | 11 (0.1)         | 29 (0.1)         |
| Unknown                   | 290 (2.7)        | 885 (2.1)        |
| **DSH risk score, mean (SD)** |                  |                  |
| Type 2 diabetes           |                  |                  |
| Sum                       | 5.4 (13.6)       | 12.2 (17.6)      |
| Mean                      | 0.4 (0.9)        | 0.9 (1.3)        |
| Last observed             | 0.2 (0.4)        | 0.2 (0.8)        |
| Hypertension              |                  |                  |
| Sum                       | 38.2 (22.7)      | 47.3 (28.7)      |
| Mean                      | 2.5 (1.5)        | 3.4 (2.0)        |
| Last observed             | 1.5 (1.6)        | 1.1 (2.3)        |
| Hyperlipidemia            |                  |                  |
| Sum                       | 19.5 (12.5)      | 18.4 (13.5)      |
| Mean                      | 1.3 (0.8)        | 1.3 (1.0)        |
| Last observed             | 0.7 (1.1)        | 0.6 (1.2)        |
| Obesity                   |                  |                  |
| Sum                       | 15.1 (12.1)      | 17.1 (12.0)      |
| Mean                      | 1.0 (0.8)        | 1.2 (0.9)        |
| Last observed             | 0.7 (1.1)        | 0.3 (0.8)        |
| **Comorbidities**         |                  |                  |
| Type 2 diabetes           |                  |                  |
| Most frequent             | 1733 (16.2)      | 15 716 (38.1)    |
| Rolling                   | 2664 (25.0)      | 26 438 (64.0)    |
| Hypertension              |                  |                  |
| Most frequent             | 10 019 (93.9)    | 29 621 (71.7)    |
| Rolling                   | 10 575 (99.1)    | 41 218 (99.8)    |
| Hyperlipidemia            |                  |                  |
| Most frequent             | 7678 (71.9)      | 21 839 (52.9)    |
| Rolling                   | 8806 (82.5)      | 28 563 (69.2)    |
| Obesity                   |                  |                  |
| Most frequent             | 4078 (38.2)      | 17 111 (41.4)    |
| Rolling                   | 4986 (46.7)      | 26 198 (63.4)    |
| **Medication classes**    |                  |                  |
| Frequency, mean (SD), No. | 23.2 (22.3)      | 14.0 (0.2)       |
| ACE inhibitor             | 3416 (32.0)      | 25 968 (62.9)    |
| Calcium channel blocker   | 2192 (20.5)      | 19 214 (46.5)    |
| β-blocker                 | 3958 (37.1)      | 30 437 (73.7)    |
| Diuretic                  | 1124 (10.5)      | 28 046 (67.9)    |
| Statin                    | 5687 (53.3)      | 31 294 (75.8)    |
| α-blocker                 | 1183 (11.1)      | 90 (0.2)         |
| Angiotensin receptor blocker | 1515 (14.2)    | 11 832 (28.6)    |
| Other                     | 262 (2.5)        | 12 440 (30.1)    |

(continued)
Table 1. Study Population (continued)

| Variable | No. (%) | REP (n = 10,674) | FHS (n = 41,295) |
|----------|---------|------------------|------------------|
| Fibrate  | 478 (4.5) | 4095.00 (9.92)  |                  |
| Sulfonylurea | 660 (6.2) | 7476 (18.1)     |                  |
| Renin inhibitor | 7 (0.1) | 70 (0.3)        |                  |
| Insulin  | 545 (5.1) | 14,559 (35.2)   |                  |
| Cholesterol absorption inhibitor | 587 (5.5) | 4213 (10.2) |                  |
| Metformin | 1196 (11.2) | 13,033 (31.6)  |                  |
| Dipeptidyl peptidase-4 | 124 (1.2) | 2150 (5.2)     |                  |
| Meglitinide | 23 (0.2) | 203 (0.5)       |                  |
| Vasodilator | 63 (0.6) | 16,619 (40.2)  |                  |
| GLP-1 agonist | 48 (0.5) | 2070 (5.0)     |                  |
| Amylin   | 4 (<0.1) | 52 (0.1)        |                  |
| SGLT-2 inhibitor | 1 (<0.1) | 608 (1.5)      |                  |

Frequency of laboratory results and vital signs, mean (SD)

| Glycated hemoglobin | 2.9 (6.4) | 7.3 (9.3) |
| BNP         | 0.2 (1.2) | 0.0 (0.0) |
| Creatinine  | 18.9 (28.5) | 29.4 (39.1) |
| Cardiac troponin 1 | 0.1 (0.5) | 0.0 (0.0) |
| Cardiac troponin T | 2.5 (6.1) | 0.0 (0.0) |
| Glucose | |
| Fasting  | 10.1 (11.3) | 28.1 (38.1) |
| Random   | 0.1 (0.4) | 33.6 (103.2) |
| Glomerular filtration rate | 11.0 (27.5) | 39.6 (66.8) |
| HDL cholesterol | 8.9 (6.8) | 8.9 (6.9) |
| LDL cholesterol | 8.7 (6.7) | 8.8 (6.8) |
| Total cholesterol | 9.0 (6.9) | 8.9 (6.9) |
| NT-proBNP | 0.3 (1.6) | 0.0 (0.0) |
| Triglycerides | 8.8 (6.9) | 9.0 (7.5) |
| BMI | 18.2 (15.5) | 38.3 (31.0) |
| Diastolic blood pressure | 106.9 (204.6) | 46.7 (39.5) |
| Height | 20.4 (18.1) | 0.0 (0.0) |
| Pulse | 196.0 (618.4) | 41.7 (35.9) |
| Systolic blood pressure | 106.9 (204.5) | 46.7 (39.5) |
| Weight | 40.2 (41.3) | 0.0 (0.0) |
| Respiration | 61.1 (222.6) | 0.0 (0.0) |

Laboratory results and vital signs, mean (SD)

| Creatinine, mg/dL | 1.0 (0.3) | 6.1 (103.2) |
| Fasting glucose, mg/dL | 104.1 (20.2) | 119.9 (32.6) |
| Total cholesterol, mg/dL | 193.3 (30.1) | 178.8 (81.8) |
| BMI | 28.4 (6.3) | 35.1 (63.9) |
| Diastolic blood pressure, mm Hg | 73.9 (7.0) | 74.0 (6.7) |
| Pulse, bpm | 73.3 (8.4) | 74.5 (8.9) |
| Systolic blood pressure, mm Hg | 127.1 (11.7) | 128.0 (10.3) |

Last observed labs and vitals, mean (SD)

| Creatinine, mg/dL | 1.0 (0.3) | 8.3 (146.9) |
| Fasting glucose, mg/dL | 103.8 (20.2) | 119.7 (33.9) |
| Total cholesterol, mg/dL | 192.1 (30.0) | 177.9 (152.3) |
| BMI | 28.3 (6.2) | 36.1 (76.8) |
| Diastolic blood pressure, mm Hg | 73.6 (7.1) | 73.8 (7.1) |
| Pulse, bpm | 73.2 (8.6) | 74.8 (9.3) |
| Systolic blood pressure, mm Hg | 126.4 (11.8) | 127.9 (10.8) |
80 years on REP data of 0.96 (95% CI, 0.94-0.97), 0.96 (95% CI, 0.95-0.98), 0.97 (95% CI, 0.96-0.98), and 0.98 (95% CI, 0.98-0.99), respectively, while models using COM produced modest AUCs of 0.67 (95% CI, 0.55-0.80), 0.66 (95% CI, 0.56-0.79), 0.64 (95% CI, 0.57-0.71), and 0.63 (95% CI, 0.54-0.70), respectively. Adding laboratory results and vital signs to COM (ie, COM + LB/VS) improved the performance of COM, resulting in AUCs between 0.73 and 0.87; however, the performance was still inferior to DSH-RS. External validation on FHS data demonstrated good performance for DSH-RS (AUCs, 0.77-0.81), outperforming COM (AUCs, 0.72-0.75) but with

| Table 1. Study Population (continued) |
|---------------------------------------|

| Variable | No. (%) | REP (n = 10,674) | FHS (n = 41,295) |
|----------|---------|----------------|-----------------|
| ACM      | 945 (8.9) | 1857 (4.5) |
| MCE      | 787 (7.4) | 3178 (7.7) |

Abbreviations: ACE, angiotensin-converting enzyme; ACM, all-cause mortality; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); BNP, brain natriuretic peptide; bpm, beats per minute; DSH, disease severity hierarchy; FHS, Fairview Health Services; GLP-1, glucagon-like peptide 1; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MCE, major cardiovascular event; NT-proBNP, N-terminal pro-brain natriuretic peptide; REP, Rochester Epidemiology Project; SGLT-2, sodium-glucose cotransporter-2.

SI conversion factors: To convert creatinine to micromoles per liter, multiply by 88.4; fasting glucose to millimoles per liter, multiply by 0.055; and total cholesterol to millimoles per liter, multiply by 0.0259.

* The index date for patients in REP was defined as age on January 1, 2004, and age on January 1, 2010, for FSH.

| Table 2. All-Cause Mortality Internal and External Validation |
|-------------------------------------------------------------|

| Model | Age, y | ACC | AUC | Sensitivity | Specificity | PPV |
|-------|--------|-----|-----|-------------|-------------|-----|
| Internal validation with cross-validation on REP: predictors in 2004-2010, follow-up in 2010-2015, mean (95% CI) |
| DSH-RS | 60 | 0.91 (0.90-0.93) | 0.96 (0.94-0.97) | 0.98 (0.91-1.00) | 0.91 (0.90-0.93) | 0.10 (0.07-0.13) |
| 65 | 0.92 (0.90-0.93) | 0.96 (0.95-0.98) | 0.99 (0.94-1.00) | 0.91 (0.90-0.93) | 0.15 (0.10-0.19) |
| 75 | 0.94 (0.93-0.95) | 0.97 (0.96-0.98) | 0.98 (0.93-1.00) | 0.94 (0.93-0.95) | 0.35 (0.25-0.39) |
| 80 | 0.96 (0.95-0.97) | 0.98 (0.98-0.99) | 0.99 (0.97-1.00) | 0.95 (0.95-0.96) | 0.55 (0.49-0.62) |
| COM | 60 | 0.62 (0.22-0.94) | 0.67 (0.55-0.80) | 0.49 (0.11-0.89) | 0.62 (0.21-0.94) | 0.02 (0.01-0.05) |
| 65 | 0.61 (0.38-0.93) | 0.66 (0.56-0.79) | 0.55 (0.17-0.85) | 0.61 (0.38-0.94) | 0.03 (0.01-0.05) |
| 75 | 0.54 (0.41-0.65) | 0.64 (0.57-0.71) | 0.64 (0.46-0.79) | 0.54 (0.40-0.65) | 0.05 (0.03-0.06) |
| 80 | 0.59 (0.51-0.67) | 0.63 (0.54-0.70) | 0.60 (0.51-0.70) | 0.59 (0.50-0.67) | 0.08 (0.05-0.10) |
| COM + LB/VS | 60 | 0.80 (0.62-0.96) | 0.86 (0.79-0.92) | 0.68 (0.42-0.83) | 0.80 (0.62-0.97) | 0.06 (0.01-0.15) |
| 65 | 0.84 (0.70-0.96) | 0.87 (0.82-0.93) | 0.70 (0.48-0.87) | 0.85 (0.70-0.97) | 0.10 (0.03-0.23) |
| 75 | 0.79 (0.69-0.86) | 0.79 (0.74-0.83) | 0.64 (0.58-0.73) | 0.80 (0.69-0.87) | 0.11 (0.05-0.17) |
| 80 | 0.72 (0.60-0.80) | 0.73 (0.67-0.78) | 0.58 (0.47-0.70) | 0.73 (0.59-0.82) | 0.11 (0.07-0.15) |
| External validation on FHS: predictors in 2010-2015, follow-up in 2015-2017, mean |
| DSH-RS | 60 | 0.76 | 0.81 | 0.72 | 0.76 | 0.05 |
| 65 | 0.75 | 0.80 | 0.68 | 0.76 | 0.07 |
| 75 | 0.76 | 0.77 | 0.63 | 0.76 | 0.11 |
| 80 | 0.74 | 0.77 | 0.65 | 0.74 | 0.11 |
| COM | 60 | 0.43 | 0.75 | 0.87 | 0.43 | 0.03 |
| 65 | 0.41 | 0.74 | 0.88 | 0.39 | 0.04 |
| 75 | 0.36 | 0.73 | 0.90 | 0.33 | 0.06 |
| 80 | 0.38 | 0.72 | 0.88 | 0.36 | 0.06 |
| COM + LB/VS | 60 | 0.42 | 0.82 | 0.94 | 0.41 | 0.03 |
| 65 | 0.46 | 0.81 | 0.92 | 0.45 | 0.04 |
| 75 | 0.41 | 0.81 | 0.95 | 0.38 | 0.07 |
| 80 | 0.35 | 0.80 | 0.96 | 0.33 | 0.06 |

Abbreviations: ACC, accuracy; AUC, area under the receiving operator characteristic curve; COM, comorbidities with medications; COM + LB/VS, COM with laboratory results and vital signs; DSH-RS, disease severity hierarchy with risk scores only; FHS, Fairview Health Services; PPV, positive predictive value; REP, Rochester Epidemiology Project.

80 years on REP data of 0.96 (95% CI, 0.94-0.97), 0.96 (95% CI, 0.95-0.98), 0.97 (95% CI, 0.96-0.98), and 0.98 (95% CI, 0.95-0.98), respectively, while models using COM produced modest AUCs of 0.67 (95% CI, 0.55-0.80), 0.66 (95% CI, 0.56-0.79), 0.64 (95% CI, 0.57-0.71), and 0.63 (95% CI, 0.54-0.70), respectively. Adding laboratory results and vital signs to COM (ie, COM + LB/VS) improved the performance of COM, resulting in AUCs between 0.73 and 0.87; however, the performance was still inferior to DSH-RS. External validation on FHS data demonstrated good performance for DSH-RS (AUCs, 0.77-0.81), outperforming COM (AUCs, 0.72-0.75) but with
approximately the same performance as COM + LB/VS (AUCs, 0.80-0.82). A similar performance trend was observed for accuracy, sensitivity, specificity, and PPV (eg, age 80 years on REP for DSH-RS: accuracy, 0.96; 95% CI, 0.95-0.97; sensitivity, 0.99; 95% CI, 0.97-1.00; specificity, 0.95; 95% CI, 0.95-0.96; PPV, 0.08; 95% CI, 0.05-0.10; for COM + LB/VS: accuracy, 0.72; 95% CI, 0.60-0.80; sensitivity, 0.58; 95% CI, 0.47-0.70; specificity, 0.59; 95% CI, 0.50-0.67; PPV, 0.55; 95% CI, 0.49-0.62; for COM: accuracy, 0.59; 95% CI, 0.51-0.67; sensitivity, 0.60; 95% CI, 0.51-0.70; specificity, 0.59; 95% CI, 0.50-0.67; PPV, 0.08; 95% CI, 0.05-0.10; for COM + LB/VS: accuracy, 0.72; 95% CI, 0.60-0.80; sensitivity, 0.58; 95% CI, 0.47-0.70; specificity, 0.73; 95% CI, 0.59-0.82; PPV, 0.11; 95% CI, 0.07-0.15; age 80 years on FHS for DSH-RS: accuracy, 0.74; sensitivity, 0.65; specificity, 0.74; PPV, 0.11; for COM: accuracy, 0.38; sensitivity, 0.88; specificity, 0.36; PPV, 0.06; for COM + LB/VS: accuracy, 0.35; sensitivity, 0.96; specificity, 0.33; PPV, 0.06).

The eTable in the Supplement shows the results for predicting MCE on REP and FHS, in which we observed a similar strong performance of DSH-RS (internal AUCs, 0.75-0.79; external AUCs, 0.65-0.67), outperforming COM (internal AUCs, 0.54-0.58; external AUCs, 0.59-0.61) and COM + LB/VS (internal AUCs, 0.58-0.75; external AUCs, 0.57-0.59) on both the internal and external validation data sets.

**Figure 3** shows the receiver operating characteristic, cumulative gain, calibration, decision, prediction density, sensitivity, and PPV performance plots for predicting ACM at age 75 years for REP data. The AUC and Gini scores for DSH-RS (0.97 and 0.95, respectively) were considerably higher than those for COM (0.65 and 0.31, respectively) and COM + LB/VS (0.79 and 0.58, respectively) (Figure 3A and Figure 3B). The gain curve (Figure 3B) shows that if we were to select the top 10% of the entire population, representing 1067 patients, as high-risk for ACM at age 75 years based on DSH-RS, the sample will contain approximately 100% of high-risk patients. On the other hand, COM and COM + LB/VS will contain only 30.0% and 55.0% of high-risk cases, respectively. Similarly, the...

---

**Figure 3. Internal Validation Performance Plots for Cox Proportional Hazard Models Predicting ACM at Age 75 Years**

A, ROC

B, Cumulative gain curve

C, Decision curve

D, Calibration

E, Prediction density: DSH-RS

F, Prediction density: COM

G, Sensitivity

H, Positive predictive value

B, The Gini score was computed by dividing the area between the gain curve and the random classifier (indicated by the dotted diagonal line) by the area between the perfect classifier (indicated by the purple curve) and the random classifier. D, the dotted diagonal line represents the line of perfect calibration. Systematic deviation (below or above) from the diagonal line indicates that the model might not reliably estimate event rates, leading to overestimation and underestimation. COM indicates model with only comorbidities and medication; COM + LB/VS, COM with laboratory results and vital signs; DSH-RS, disease severity hierarchy-risk score; ROC, receiver operating characteristic curve.
net benefit (Figure 3C), sensitivity (Figure 3G), and PPV (Figure 3H) curves of DSH-RS were higher than those for COM and COM + LB/VS. The class distributions in the prediction density plots (Figure 3E and Figure 3F), in which yes indicates death by age 75 years and no indicates survival older than 75 years, showed that DSH-RS accurately discriminated between low-risk and high-risk patients compared with COM. Complete graphic performance plots for ACE and MCE for REP data are available in eFigures 4 to 10 in the Supplement.

Discussion

In this study, we introduced DSH, a novel representation of patient data that explores diseases and their known treatment pathways in a nested fashion. Unlike conventional methods of representing comorbidities as independent, time-invariant, and simple summaries, DSH offers explicit relationships for progression and severity of comorbidities as they relate to health status. We showed that these associations between comorbidities and their severity inform clinical risk stratification. This simple and expressive yet succinct representation of health status can be easily introduced to many clinical analysis processes and has the potential to greatly improve predictions. We demonstrated this capability by predicting age-time of ACM and MCE based on DSH representation of type 2 diabetes and its comorbidities. Models based on DSH representations outperformed models based on individual comorbidities, medications, laboratory results, and vital signs.

In the era of value-based health care, when practitioners are paid based on patient health outcomes, efficient risk stratification and patient management are becoming more important than ever. Health care organizations working to change their cost structure and improve outcomes must design interventions that target high-risk patients who need to be managed carefully and proactively. The first step to targeting high-risk patients is to identify them. Thus, implementing a platform to accurately stratify patients according to risk is the cornerstone of success for any population health management initiative.27 Several risk stratification systems, such as the Chronic Comorbidity Count, Adjusted Clinical Groups, Hierarchical Condition Categories, the Charlson-Elixhauser Comorbidity Index, and the Minnesota Health Care Home Tiering, among others, have been proposed.28,29 A commonality among these models is that they are based in some degree on comorbidity. This is not surprising given that understanding comorbid conditions and the heterogeneity exhibited by patients with MCCs is the first step in implementing effective population health management initiatives.27-29 Unfortunately, current representations of comorbidity indices do not reflect an individual’s comorbid disease history and severity or account for interactions with different treatment pathways.

This study was heavily driven by clinical domain knowledge and inspired by clinical thinking created by clinicians closely collaborating with data scientists. We aimed to make use of known relationships about specific diseases and their management and outcomes to improve the performance of machine learning models. By incorporating a priori domain knowledge into the data representations, we reduced the complexity of the problem, allowing the model to put observations in the clinical context, which otherwise would not be included in the modeling task. By explicitly modeling disease severity and treatment pathways, DSH-based models can improve human interpretability compared with competing methods.

Our ultimate goal was to generate information that can be used to develop EBPGs implementable in clinical decision support systems. EBPGs are series of interventions, each intervention being defined by a clinical action and a subpopulation for which this action would be most beneficial. Objective and accurate risk assessment tools are critically important for developing smart EBPGs and for counselling patients regarding the potential benefits and risks of different treatment strategies. The proposed DSH aimed to precisely identify patient subpopulations for which interventions may be useful.26
Limitations
This study has limitations. First, the manual construction of DSH can be error prone and time consuming. How to automate the process, while remaining interpretable, is an open research question. Second, DSH uses retrospective EHR data, in which treatment pathways and disease intensity were determined based on whether a treatment was given. However, EHR data contain errors, and an erroneous treatment assignment may lead to inaccurate encoding of disease severity. Third, we have not investigated the generalizability of our methods to other complex comorbidities, nor have we investigated other functional relationships between the DSH nodes, other than doubling the score on branches to the right. It is not clear whether the results would remain consistent if other relationships (e.g., an exponential relationship) were used. These are important open research questions to be investigated.

Conclusions
In this study, we developed a novel data representation scheme and validated the added value in a risk prediction model. The model accurately predicted the age at which a patient is at risk of ACM or MCE substantially better than standard data presentations. DSH uses expert knowledge of known relationships within EHR data to capture disease severity in a natural and clinically meaningful way. The proposed DSH and corresponding risk scoring system can help to support critical decision-making, develop smart EBPGs, and enhance health care and disease management. It also has the potential to be used as a health care quality metric.

ARTICLE INFORMATION
Accepted for Publication: April 13, 2020.
Published: July 17, 2020. doi:10.1001/jamanetworkopen.2020.8270
Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2020 Ngufor C et al. JAMA Network Open.

Corresponding Author: Che Ngufor, PhD, Division of Digital Health Sciences, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905 (ngufor.che@mayo.edu).

Author Affiliations: Division of Digital Health Science, Department of Health Science Research, Mayo Clinic, Rochester, Minnesota (Ngufor, Caraballo, Chen); The Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota (Ngufor, Shah); Division of General Internal Medicine, Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota (Caraballo, Simon); Division of Healthcare Policy and Research, Department of Health Science Research, Mayo Clinic, Rochester, Minnesota (O’Byrne, Shah); University of Minnesota School of Nursing, Minneapolis (Pruinelli); Department of Computer Science and Engineering, University of Minnesota, Minneapolis (Steinbach); Institute for Health Informatics, University of Minnesota, Minneapolis (Simon); Department of Medicine, University of Minnesota, Minneapolis (Simon).

Author Contributions: Drs Ngufor and Simon had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Ngufor, Caraballo, Steinbach, Simon.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Ngufor, Caraballo, Simon.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Ngufor, O’Byrne, Chen, Simon.
Obtained funding: Caraballo, Steinbach, Simon.
Administrative, technical, or material support: Ngufor, Caraballo.
Supervision: Caraballo, Steinbach, Simon.

Conflict of Interest Disclosures: Dr Ngufor reported receiving grants and nonfinancial support consisting of office space and supplies, computers, and other administrative services from the Mayo Clinic and the University of Minnesota during the conduct of the study. Dr Shah reported receiving research support through the Mayo Clinic.
from the Centers of Medicare & Medicaid Innovation, the US Food and Drug Administration, the Agency for Healthcare Research and Quality, the National Heart, Lung, and Blood Institute of the National Institutes of Health, the Medical Devices Innovation Consortium/National Evaluation System for Health Technology, the National Science Foundation, and the Patient-Centered Outcomes Research Institute. No other disclosures were reported.

**Funding/Support:** This research was supported by grants IIS-1602198 and IIS-1602394 from the National Science Foundation and grant LMI1972 from the National Institutes of Health.

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Disclaimer:** The views expressed in this manuscript are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

**REFERENCES**

1. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. *Prev Chronic Dis*. 2014;11:E62. doi:10.5888/pcd11.130389

2. Freid VM, Bernstein AB, Bush MA. Multiple chronic conditions among adults aged 45 and over: trends over the past 10 years. *NCHS Data Brief*. 2012;(100):1-8.

3. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med*. 2002;162(20):2269-2276. doi:10.1001/archinte.162.20.2269

4. Sambamortu OI, Tan X, Deb A. Multiple chronic conditions and healthcare costs among adults. *Expert Rev Pharmacoecon Outcomes Res*. 2015;15(5):823-832. doi:10.1586/14737167.2015.1091730

5. Lehnert T, Heider D, Leicht H, et al. Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev*. 2011;68(4):387-420. doi:10.1177/1077558711399580

6. Chu Y-T, Ng Y-Y, Wu S-C. Comparison of different comorbidity measures for use with administrative data in predicting short- and long-term mortality. *BMJ Health Serv Res*. 2010;10(1):140. doi:10.1186/1472-6963-10-140

7. Vogeli C, Shields AE, Lee TA, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *J Gen Intern Med*. 2007;22(3)(suppl 3):391-395. doi:10.1007/s11606-007-0322-1

8. Equator Network. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Accessed March 31, 2020. https://www.equator-network.org/reporting-guidelines/strobe/

9. Melton LJ III. History of the Rochester Epidemiology Project. *Mayo Clin Proc*. 1996;71(3):266-274. doi:10.4065/71.3.266

10. Ramalho de Oliveira D, Brummel AR, Miller DB. Medication therapy management: 10 years of experience in a large integrated health care system. *J Manag Care Pharm*. 2010;16(3):185-195. doi:10.18553/jmcp.2010.16.3.185

11. Griffin BA, Anderson GL, Shih RA, Whitsel EA. Use of alternative time scales in Cox proportional hazard models: implications for time-varying environmental exposures. *Stat Med*. 2012;31(27):3320-3327. doi:10.1002/sim.5347

12. Lamarca R, Alonso J, Gómez G, Muñoz A. Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. *J Gerontol A Biol Sci Med Sci*. 1998;53(5):M337-M343. doi:10.1093/gerona/53A.5.M337

13. Canchola AJ, Stewart SL, Bernstein L, et al. *Cox Regression Using Different Time-Scales*. Western Users of SAS Software San Francisco; 2003.

14. Association AD; American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2019. *Diabetes Care*. 2019;42(suppl 1):S13-S28. doi:10.2337/dc19-S002

15. Ngufor C, Van Houten H, Caffo BS, Shah ND, McCoy RG. Mixed effect machine learning: a framework for predicting longitudinal change in hemoglobin A1c. *J Biomed Inform*. 2019;89:56-67. doi:10.1016/j.jbi.2018.09.001

16. Austin PC. A tutorial on multilevel survival analysis: methods, models and applications. *Int Stat Rev*. 2017;85(2):185-203. doi:10.1111/insr.12214

17. Wang CY, Baldwin L-M, Saver BG, et al. The contribution of longitudinal comorbidity measurements to survival analysis. *Med Care*. 2009;47(7):813-821. doi:10.1097/MLR.0b013e318197929c

18. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

19. Ko Y, Cho M, Lee J-S, Kim J. Identification of disease comorbidity through hidden molecular mechanisms. *Sci Rep*. 2016;6:39433. doi:10.1038/srep39433

20. Ko Y, Cho M, Lee J-S, Kim J. Identification of disease comorbidity through hidden molecular mechanisms. *Sci Rep*. 2016;6:39433. doi:10.1038/srep39433
20. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw.* 2011;39(5):1-13. doi:10.18637/jss.v039.i05

21. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc B.* 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

22. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics.* 1970;12(1):55-67. doi:10.1080/00401706.1970.10488634

23. Greene HJ, Milne GR. Assessing model performance: the Gini statistic and its standard error. *J Database Marketing Customer Strategy Manage.* 2010;17(1):36-48. doi:10.1057/dbm.2010.2

24. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol.* 2017;17(1):53. doi:10.1186/s12874-017-0332-6

25. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* 2008;8(1):53. doi:10.1186/1472-6947-8-53

26. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ.* 2016;352:i6. doi:10.1136/bmj.i6

27. Just E. Understanding risk stratification, comorbidities, and the future of healthcare. Published November 25, 2016. Accessed June 15, 2020. https://www.healthcatalyst.com/understanding-risk-stratification-comorbidities/

28. Naessens JM, Stroebel RJ, Finnie DM, et al. Effect of multiple chronic conditions among working-age adults. *Am J Manag Care.* 2011;17(2):118-122.

29. Haas LR, Takahashi PY, Shah ND, et al. Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care.* 2013;19(9):725-732.

**SUPPLEMENT.**

eAppendix. *International Classification of Diseases, Ninth Revision (ICD-9) and ICD-10-CM codes*
eTable. *Major Cardiovascular Event Internal and External Validation*
eFigure 1. Associations Between Missingness and Observations in Rochester Epidemiology Data
eFigure 2. Associations Between Missingness and Observations in Fairview Health System Data
eFigure 3. Disease Severity Hierarchy for Obesity
eFigure 4. Predicted All-Cause Mortality at Age 60 Years
eFigure 5. Predicted All-Cause Mortality at Age 65 Years
eFigure 6. Predicted All-Cause Mortality at Age 80 Years
eFigure 7. Predicted Major Cardiovascular Event at Age 60 Years
eFigure 8. Predicted Major Cardiovascular Event at Age 65 Years
eFigure 9. Predicted Major Cardiovascular Event at Age 75 Years
eFigure 10. Predicted Major Cardiovascular Event at Age 80 Years