ORIGINAL RESEARCH

Angina Severity, Mortality, and Healthcare Utilization Among Veterans With Stable Angina

Mina Owlia, MD; John A. Dodson, MD; Jordan B. King, PharmD, MS; Catherine G. Derington, PharmD, MS; Jennifer S. Herrick, MS; Steven P. Sedlis, MD; Jacob Crook, MSTAT; Scott L. DuVall, PhD; Joanne LaFleur, PharmD, MSPH; Richard Nelson, PhD; Olga V. Patterson, PhD; Rashmee U. Shah, MD, MS; Adam P. Bress, PharmD, MS

***Background*—**Canadian Cardiovascular Society (CCS) angina severity classification is associated with mortality, myocardial infarction, and coronary revascularization in clinical trial and registry data. The objective of this study was to determine associations between CCS class and all-cause mortality and healthcare utilization, using natural language processing to extract CCS classifications from clinical notes.

***Methods and Results*—**In this retrospective cohort study of veterans in the United States with stable angina from January 1, 2006, to December 31, 2013, natural language processing extracted CCS classifications. Veterans with a prior diagnosis of coronary artery disease were excluded. Outcomes included all-cause mortality (primary), all-cause and cardiovascular-specific hospitalizations, coronary revascularization, and 1-year healthcare costs. Of 299 577 veterans identified, 14 216 (4.7%) had ≥1 CCS classification extracted by natural language processing. The mean age was 66.6±9.8 years, 99% of participants were male, and 81% were white. During a median follow-up of 3.4 years, all-cause mortality rates were 4.58, 4.60, 6.22, and 6.83 per 100 person-years for CCS classes I, II, III, and IV, respectively. Multivariable adjusted hazard ratios for all-cause mortality comparing CCS II, III, and IV with those in class I were 1.05 (95% CI, 0.95–1.15), 1.33 (95% CI, 1.20–1.47), and 1.48 (95% CI, 1.25–1.76), respectively. The multivariable hazard ratio comparing CCS IV with CCS I was 1.20 (95% CI, 1.09–1.33) for all-cause hospitalization, 1.25 (95% CI, 0.96–1.64) for acute coronary syndrome hospitalizations, 1.00 (95% CI, 0.80–1.26) for heart failure hospitalizations, 1.05 (95% CI, 0.88–1.25) for atrial fibrillation hospitalizations, 1.92 (95% CI, 1.40–2.64) for percutaneous coronary intervention, and 2.51 (95% CI, 1.99–3.16) for coronary artery bypass grafting surgery.

***Conclusions*—**Natural language processing–extracted CCS classification was positively associated with all-cause mortality and healthcare utilization, demonstrating the prognostic importance of anginal symptom assessment and documentation. **(*J Am Heart Assoc.* 2019;8:e012811. DOI: 10.1161/JAHA.119.012811.)**

Key Words: angina pectoris • healthcare utilization • myocardial revascularization • natural language processing

Stable angina affects >10 million Americans and is the presenting symptom in approximately half of patients with coronary artery disease.[1–3] Contemporary evidence-based interventions for stable angina include aggressive lifestyle modifications, pharmacotherapy, and coronary revascularization to improve anginal symptoms and overall health status.[2] An accurate, comprehensive assessment of symptom severity at the point of care is difficult to estimate but holds important implications for shared medical decision-making between patient and clinician, who work together to determine the specific types and intensity of interventions for angina treatment based on symptomatology.

Stable angina severity can be assessed in clinical practice and research settings by the physician using grading

From the Leon H. Charney Division of Cardiology, Departments of Medicine (M.O., J.A.D., S.P.S.) and Population Health (J.A.D.), New York University School of Medicine, New York, NY; Departments of Population Health Sciences (J.B.K., J.S.H., A.P.B.), Internal Medicine (J.C., S.L.D., R.N., O.V.P., R.U.S., A.P.B.), and Pharmacotherapy (J.L.), University of Utah, Salt Lake City, UT; Department of Pharmacy, Kaiser Permanente Colorado, Aurora, CO (J.B.K., C.G.D.); Department of Clinical Pharmacy, University of Colorado Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO (C.G.D.); Cardiology Section, Veterans Affairs Medical Center, Manhattan, NY (S.P.S.); George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT (J.B.K., J.C., S.L.D., J.L., R.N., O.V.P., A.P.B.).

Accompanying Data S1 and Tables S1 through S4 are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.119.012811

Correspondence to: Adam P. Bress, PharmD, MS, University of Utah School of Medicine, 295 Chipeta Way, Salt Lake City, UT 84112-5820. E-mail: adam.bress@hsc.utah.edu

Received March 29, 2019; accepted June 14, 2019.

Clinical Perspective

What Is New?

- Canadian Cardiovascular Society (CCS) anginal symptom class extracted from clinical notes within the electronic health record was positively associated with higher rates of mortality, cardiovascular hospitalizations, coronary revascularization, and overall healthcare costs.

What Are the Clinical Implications?

- Assessing and documenting anginal symptom burden at the point of care using the CCS classification system has prognostic implications.
- Our findings underscore the value of documenting CCS classification for risk stratification, shared decision-making, and allocation of resources to patient populations that would benefit the most from revascularization interventions or other medical measures.

measures such as the Canadian Cardiovascular Society (CCS) angina classification system. The CCS angina classification is a physician-reported symptom severity scale used to assess and grade physical-activity symptoms on 4 levels: class I indicates angina with strenuous exertion; class II indicates angina with walking >200 yards on flat surfaces, climbing stairs rapidly, or in cold or emotional situations; class III indicates angina with walking 100–200 yards on flat surfaces; and class IV indicates angina at rest or with any physical activity.[2,4] CCS angina classification has been associated with coronary revascularization, myocardial infarction, cognitive impairment, and mortality in clinical trials and prospective registries.[5–7] However, these associations have not been replicated in large, population-based cohorts using electronic health record data sets. This is predominantly because a patient's CCS class is documented as unstructured free text within clinic notes and thus is not easily extractable for research purposes. As such, data demonstrating the importance of CCS classification and outcomes in large cohorts of community-treated angina patients is limited, particularly in patients with newly diagnosed stable angina.

We undertook this study with 2 objectives: (1) identify CCS documentation in clinical notes within a large, integrated health system using natural language processing (NLP) techniques and (2) determine the association between initial CCS classification and all-cause mortality and healthcare utilization. We hypothesized that an NLP algorithm could accurately identify CCS classification within a large electronic health record system and that higher CCS classification would be associated with progressively higher all-cause mortality and healthcare utilization, independent of other baseline characteristics.

Methods

Study Design and Population

We conducted a retrospective cohort study using the Veterans Health Administration (VHA) clinical and administrative databases in 2 phases. First, to develop the NLP tool, we identified veterans with at least 1 inpatient or outpatient encounter with an *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) code of 413.x (angina pectoris) between January 1, 1999, and December 31, 2006. Next, we used a combination of pharmacy fill data, Current Procedural Terminology, and *ICD-9-CM* codes to identify veterans with a diagnosis of stable angina using the protocol adapted from Phelps et al (Table S1).[8] We used this cohort to develop and validate the CCS classification NLP tool (see "Natural Language Processing"). In the second phase, we employed this NLP tool, described in detail below, to identify documented CCS classifications in narrative, free-text clinical notes (eg, medical notes from office visits). We further restricted the cohort to veterans with a new diagnosis of stable angina and at least 1 CCS classification documented in clinical notes between January 1, 2006, and December 31, 2013. The index date was the first date of documentation of CCS classification for each veteran. Veterans were followed from the index date until the occurrence of the event of interest, death, or December 31, 2013, whichever occurred first.

To capture the association of CCS class at initial diagnosis with the study outcomes, the analysis was restricted to veterans without evidence of a prior diagnosis of coronary artery disease or angina using the Phelps et al criteria in the previous 1 year from the time of the first CCS class documentation (ie, patients were recently asymptomatic, not newly symptomatic; Figure 1).[8] This technique has been shown to be an effective method for identifying a cohort free of coronary artery disease using a 1-year look-back period.[9] One year was chosen because we sought to include veterans who were relatively close to their diagnosis of coronary artery disease to avoid biasing the primary outcome of all-cause mortality.

The University of Utah institutional review board approved this study with a waiver of informed consent, and the Salt Lake City Veterans Affairs Health Care System Research and Development Office approved this study. The data that support the findings of this study are available from the corresponding author on reasonable request and with appropriate Veterans Administration institutional review board and data use agreement approvals.

Data Sources

The VHA databases serve as repositories for clinical, pharmacy, and administrative data from >140 VHA

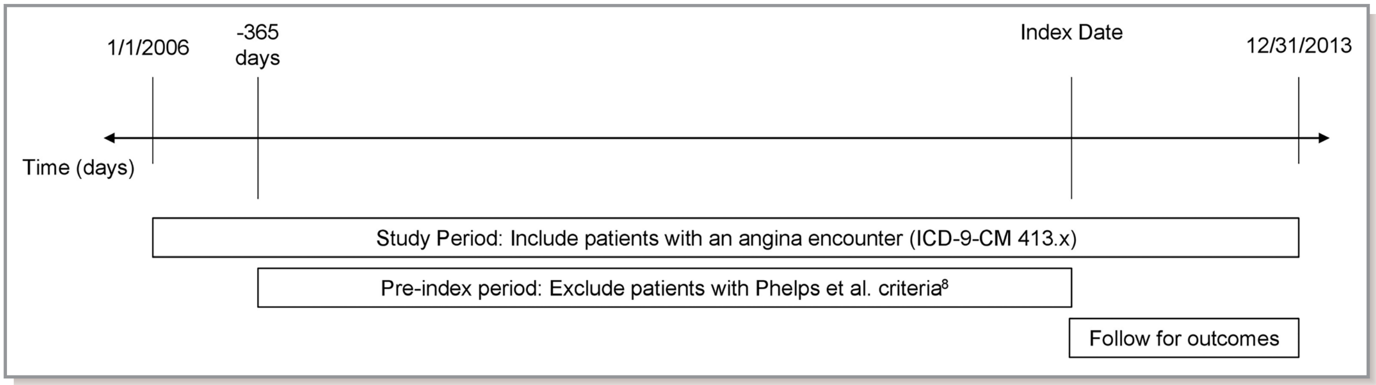


Figure 1. Application of inclusion and exclusion criteria relative to index date. *ICD-9* indicates *International Classification of Diseases, Ninth Revision*.

hospitals and 1200 outpatient clinics in all 50 states, the District of Columbia, and unincorporated territories (Guam, American Samoa, Puerto Rico, and the US Virgin Islands), representing the largest integrated healthcare delivery network in the United States. We obtained demographic, clinical, and healthcare utilization data from the Corporate Data Warehouse. For the primary outcome of all-cause mortality, the Vital Status file was used. Pharmacy data were obtained from Managerial Cost Accounting, Pharmacy Benefits Management, and the Corporate Data Warehouse. Cost data were obtained from Managerial Cost Accounting. Notes from CART-CL (VA Cardiovascular Assessment, Reporting and Tracking System for Cath Labs), a documentation system for cardiac catheterization procedures, were not included to avoid bias toward veterans undergoing revascularization procedures. Each veteran was assigned a unique deidentified number to link data sets and data tables.

Natural Language Processing

A rule-based information extraction tool was developed to identify mentions of CCS classifications and to extract the values from free-text clinical notes (Figure 2 and Table S2). The system was built using Leo, an NLP architecture developed by Veterans Informatics and Computing Infrastructure (VINCI) that uses a set of libraries that facilitate rapid creation and deployment of Apache UIMA-AS (Unstructured Information Management Architecture Asynchronous Scaleout).[10,11] The knowledge base for the system was created using NLP-assisted annotation based on a manual bootstrapping process. Precision validation was performed using the Chex Validation tool,[12] and recall validation was performed using the eHOST application.[13,14] The system achieved 93.1% precision and 75.7% recall. See Data S1 for further detail.

Outcome Measures

The primary outcome was all-cause mortality. Secondary outcomes included all-cause hospitalizations, cardiovascular-related hospitalizations (ie, acute coronary syndrome, heart failure, and atrial fibrillation), coronary revascularization procedures (ie, percutaneous coronary intervention [PCI] and coronary artery bypass grafting surgery), and 1-year direct healthcare costs (ie, inpatient, outpatient, and pharmacy costs). Cardiovascular-related hospitalizations and revascularization procedures were determined using *ICD-9* and *ICD-10* codes. The coding algorithms and data sources used to identify the outcomes in the analysis are available in Table S3. Costs are presented in 2013 US dollars.

Covariates

Demographic variables were obtained from the Corporate Data Warehouse. Clinical characteristics and comorbidities were identified from Medical SAS data tables in VINCI using *ICD-9-CM* codes to define these covariates during the 1-year pre–index date look-back period (Table S3). We calculated a Charlson Comorbidity Index for each patient using algorithms developed by Quan et al that have been used to estimate prevalence of common comorbidities in administrative data and account for disease burden within outcomes research.[15] We constructed medication profiles for each veteran using dispensing data for medications overlapping the index date. In addition, we ascertained baseline characteristics for veterans with and without a CCS classification available. For all multivariable-adjusted regression analyses, covariates were chosen a priori and included in 2 nested models. Model 1 was adjusted for age, sex, and race/ethnicity. Model 2 included the variables in model 1 with additional adjustment for dyslipidemia, diabetes mellitus, hypertension, heart failure, smoking status, body mass

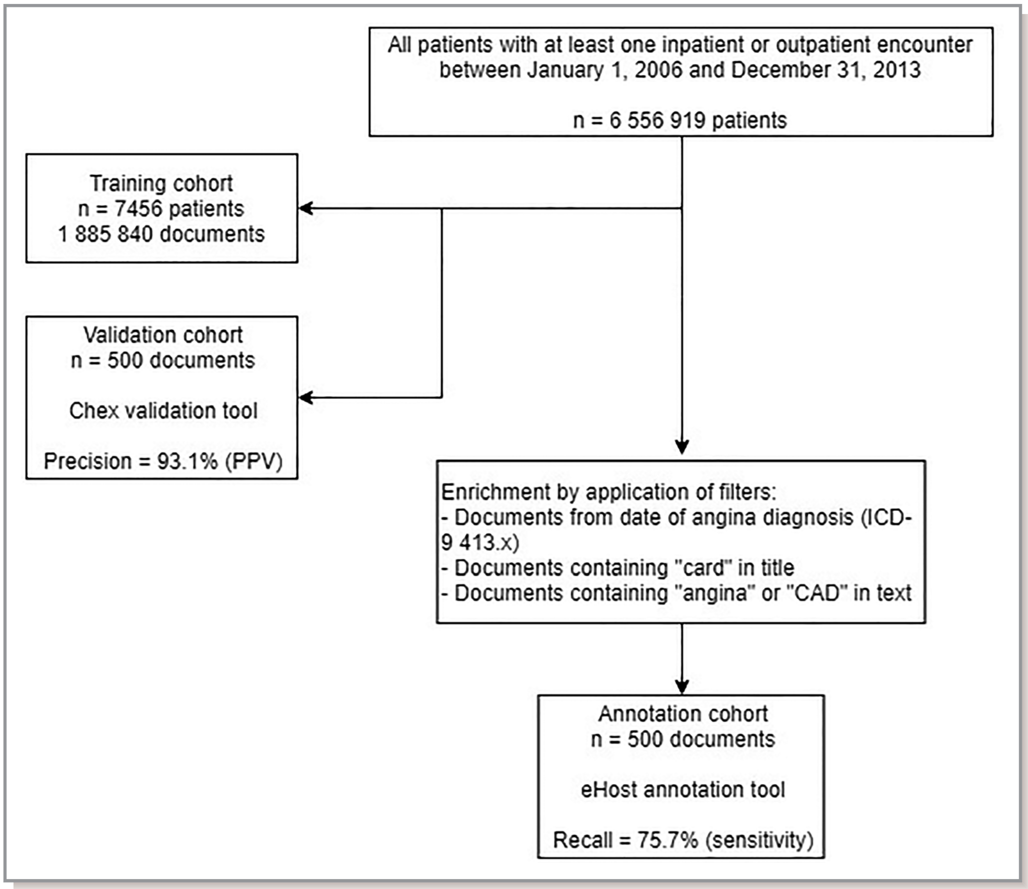


Figure 2. Development of the natural language processing tool. CAD indicates coronary artery disease; *ICD-9*, *International Classification of Diseases, Ninth Revision*; PPV, positive predictive value.

index, and medication use (aspirin, statins, β-blockers, calcium channel blockers, long-acting nitrates, angiotensin-converting enzyme inhibitors, and angiotensin receptor blockers) at baseline. The same covariates were used for cost adjustments as in models 1 and 2.

Statistical Analysis

Characteristics of the study population were calculated according to initial CCS classification ("baseline"). Trends in baseline characteristics across CCS classification were determined by modeling the characteristic of interest as the outcome variable and CCS classification as an ordinal predictor; appropriate regression models (ie, linear, logistic) were selected based on the type of variable (ie, continuous, binary) with dummy indicators created for categorical characteristics. We calculated incidence rates for each outcome as the number of events divided by the total person-years at risk. Multivariable Cox proportional hazards regression models were used to calculate hazard ratios (HRs) for all-cause mortality, hospitalizations, and revascularization procedures associated with CCS classification, with the lowest class (CCS class I) serving as the reference group. Potential violations of the proportional hazards assumption were tested by modeling the interaction between each model covariate and the log of follow-up time. No violations were present. To compare healthcare costs at 1-year follow-up between groups, adjusted incremental costs were calculated using multivariate-adjusted generalized linear models with a log-link function and a γ distribution with cost as the outcome variable and CCS class as an ordinal predictor. Finally, we compared baseline characteristics between veterans with and without CCS class measurements as identified using the NLP tool (Table S4). Analyses were performed using SAS v9.2 (SAS Institute) and STATA v12.0 (StataCorp).

Results

Patient Characteristics

Of 299 577 veterans who met criteria for a diagnosis of stable angina, 14 216 veterans (4.7%) were newly diagnosed with at least 1 documented CCS classification. The distribution of CCS classes I, II, III, and IV at baseline was 28% (n=3983), 39%

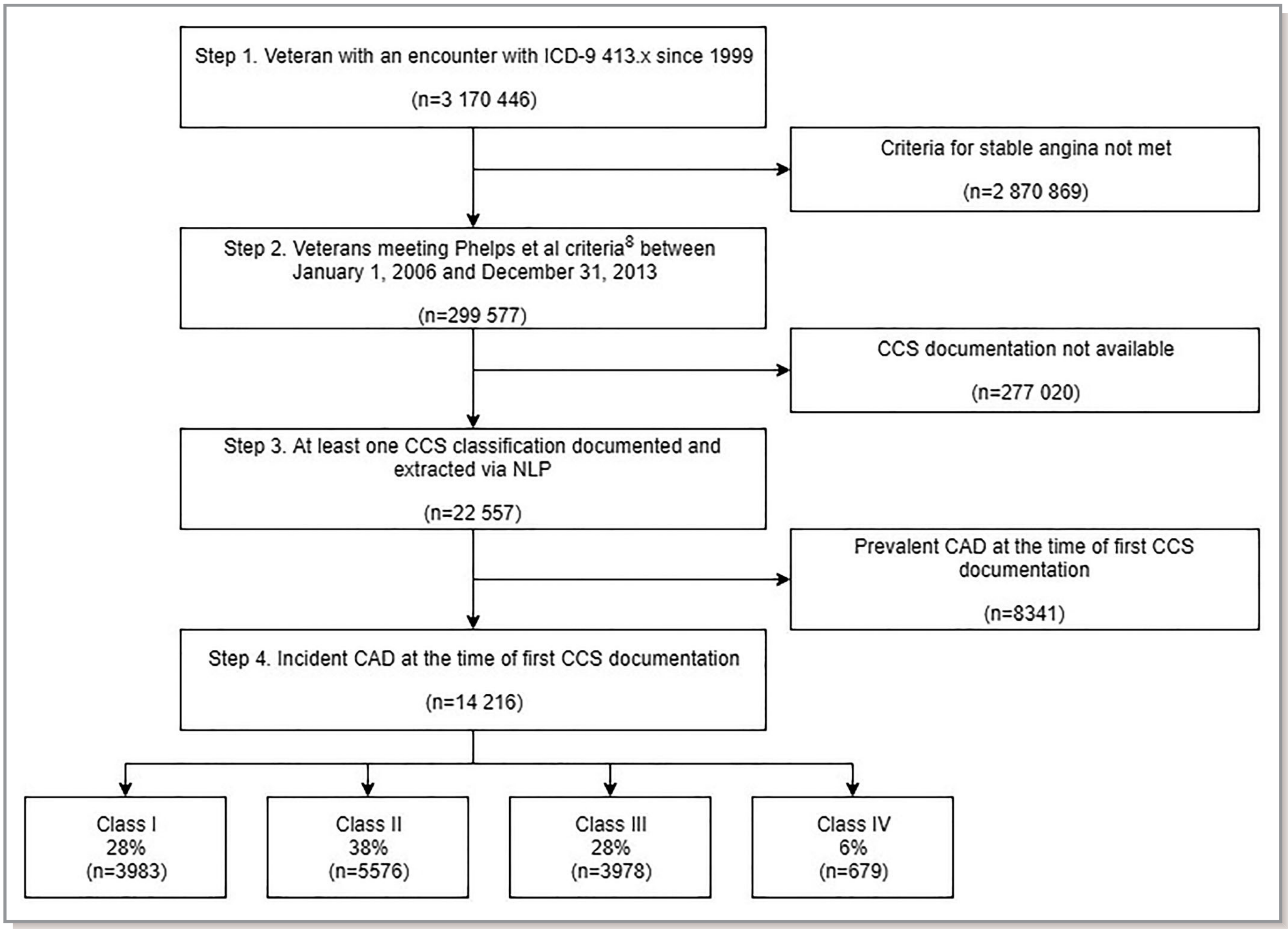


Figure 3. Flowchart of inclusion criteria defining study population. CAD indicates coronary artery disease; CCS, Canadian Cardiovascular Society (angina classification); *ICD-9*, *International Classification of Diseases, Ninth Revision*; NLP, natural language processing.

(n=5576), 28% (n=3978), and 5% (n=679), respectively (Figure 3). The mean age of the cohort was 67±9.8 years, 98% were men, and 81% were white (Table 1). Hypertension and hyperlipidemia were present in more than half the study population and did not differ between CCS classifications. Veterans with greater angina severity were more likely to have a diagnosis of diabetes mellitus, heart failure, and a higher mean Charlson comorbidity index score at baseline. Medical therapy did not vary significantly between CCS classifications at baseline except for long-acting nitrates, which were more common among veterans with CCS classes II and III.

All-Cause Mortality

During a median follow-up of 3.4 years (interquartile range: 1.6–5.6 years), all-cause mortality rates were 4.58, 4.60, 6.22, and 6.84 per 100 person-years for CCS classes I, II, III, and IV, respectively (Table 2 and Figure 4). The adjusted HRs for all-cause mortality associated with classes II, III, and IV compared with class I were 1.05 (95% CI, 0.95–1.15), 1.33 (95% CI, 1.20–1.47), and 1.48 (95% CI, 1.25–1.76), respectively ($P<0.001$ for trend).

Hospitalizations

The incidence of all-cause hospitalizations significantly increased with higher CCS class ($P<0.0001$ for trend; Table 2 and Figure 5). After full multivariable adjustment, the adjusted HRs for experiencing at least 1 all-cause hospitalization associated with classes II, III, and IV compared with class I were 1.10 (95% CI, 1.04–1.15), 1.19 (95% CI, 1.13–1.26), and 1.20 (95% CI, 1.09–1.33), respectively ($P<0.0001$ for trend). There was no association between CCS class and hospitalizations for atrial fibrillation or heart failure. Hospitalizations for acute coronary syndrome increased with CCS class, although this trend was

Table 1. Baseline Characteristics by CCS Classification

Characteristic	CCS Classification I (n=3983)	II (n=5576)	III (n=3978)	IV (n=679)	*P* Value for Trend
Age, y	67±9.8	66±9.6	67±9.9	67±10.2	0.8159
Male sex	3927 (99)	5491 (98)	3910 (98)	668 (98)	0.3312
Race					
White	3235 (81)	4469 (80)	3186 (80)	574 (85)	0.8453
Black	447 (11)	662 (12)	495 (12)	58 (9)	0.8600
Asian	74 (2)	102 (2)	50 (1)	14 (2)	0.1544
American Indian	28 (1)	61 (1)	41 (1)	5 (1)	0.3081
Unknown/missing	199 (5)	282 (5)	206 (5)	28 (4)	0.8220
Commercial insurance					
Yes	2053 (52)	2926 (52)	2095 (53)	364 (54)	0.2261
No	1242 (31)	1729 (31)	1279 (32)	212 (31)	0.4459
BMI, kg/m^2					
Underweight	25 (1)	32 (1)	32 (1)	9 (1)	0.0651
Normal	649 (16)	802 (14)	602 (15)	105 (15)	0.2501
Overweight	1483 (37)	2019 (36)	1337 (34)	244 (36)	0.0040
Obese	1826 (46)	2723 (49)	2007 (50)	321 (47)	0.0010
Comorbidities					
CCI	1.1±2.1	0.9±1.9	1.2±2.1	1.2±2.0	0.0278
Dyslipidemia	2476 (62)	3393 (61)	2376 (60)	425 (63)	0.1335
Diabetes mellitus	1743 (44)	2617 (47)	1955 (49)	335 (49)	<0.0001
Hypertension	3155 (79)	4475 (80)	3202 (80)	548 (81)	0.1443
Heart failure	770 (19)	894 (16)	875 (22)	138 (20)	0.0017
Smoking	819 (21)	1146 (21)	825 (21)	159 (23)	0.2981
Alcohol abuse	242 (6)	317 (6)	234 (6)	45 (7)	0.9261
Medications					
β-Blocker	2852 (72)	4049 (73)	2859 (72)	473 (70)	0.6501
Calcium channel blocker	934 (23)	1359 (24)	1015 (26)	159 (23)	0.1175
Long acting nitrate	1238 (31)	2230 (40)	1631 (41)	255 (38)	<0.0001
Aspirin	1362 (34)	1926 (35)	1420 (36)	221 (33)	0.5275
Statin	2857 (72)	4025 (72)	2845 (72)	469 (69)	0.3371
ACEI or ARB	2462 (62)	3408 (61)	2399 (60)	396 (58)	0.0562
$P2Y_{12}$ inhibitor	878 (22)	1232 (22)	969 (24)	147 (22)	0.0699

Data are number of patients (%) or mean (SD). ACEI indicates angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; BMI, body mass index; CCI, Charlson comorbidity index; CCS, Canadian Cardiovascular Society.

significant only in the unadjusted model (unadjusted P=0.05 for trend; fully adjusted P=0.09 for trend).

Revascularizations

A significant association was noted between increasing CCS class and risk of revascularization by either PCI or coronary artery bypass grafting (Table 3 and Figure 6). The fully adjusted HRs for PCI procedures for classes II, III, and IV relative to class I were 1.51 (95% CI, 1.25–1.83), 1.87 (95% CI, 1.53–2.27), and 1.92 (95% CI, 1.40–2.64), respectively (P<0.0001 for trend). The fully adjusted HRs for coronary artery bypass grafting for classes II, III, and IV relative to class I were 1.72 (95% CI, 1.48–1.99), 2.26 (95% CI, 1.94–2.62), and 2.51 (95% CI, 1.99–3.16), respectively (P<0.0001 for trend).

ORIGINAL RESEARCH

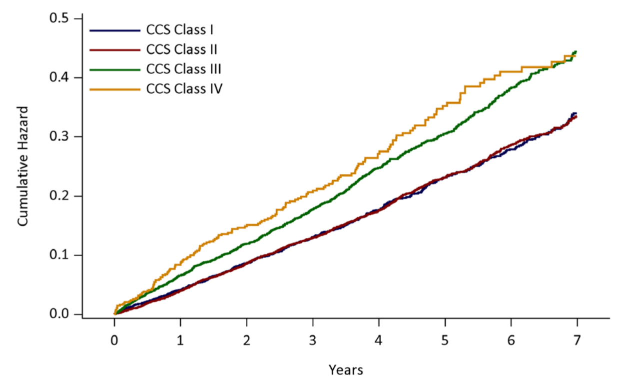
Table 2. Incidence Rates and HRs for Mortality and Hospitalizations by CCS Classification

Outcomes	CCS Classification				*P* Value for Trend
	I (n=3983)	II (n=5576)	III (n=3978)	IV (n=679)	
All-cause mortality					
Incidence rate	4.59	4.60	6.22	6.84	
HR (95% CI)					
Unadjusted	1.00 (reference)	1.00 (0.90–1.10)	1.36 (1.22–1.50)	1.49 (1.25–1.76)	<0.0001
Model 1*	1.00 (reference)	1.04 (0.94–1.15)	1.40 (1.26–1.55)	1.55 (1.31–1.84)	<0.0001
Model 2†	1.00 (reference)	1.05 (0.95–1.15)	1.33 (1.20–1.47)	1.48 (1.25–1.76)	<0.0001
All-cause hospitalization					
Incidence rate	34.56	36.96	43.26	42.51	
HR (95% CI)					
Unadjusted	1.00 (reference)	1.09 (1.03–1.14)	1.21 (1.15–1.28)	1.21 (1.09–1.33)	<0.0001
Model 1*	1.00 (reference)	1.09 (1.04–1.14)	1.21 (1.15–1.28)	1.21 (1.09–1.33)	<0.0001
Model 2†	1.00 (reference)	1.10 (1.04–1.15)	1.19 (1.13–1.26)	1.20 (1.09–1.33)	<0.0001
Acute coronary syndrome					
Incidence rate	2.24	2.35	2.56	2.87	
HR (95% CI)					
Unadjusted	1.00 (reference)	1.06 (0.92–1.23)	1.14 (0.98–1.33)	1.28 (0.98–1.67)	0.05
Model 1*	1.00 (reference)	1.08 (0.94–1.25)	1.15 (0.99–1.34)	1.31 (1.00–1.70)	0.04
Model 2†	1.00 (reference)	1.07 (0.93–1.24)	1.10 (0.95–1.29)	1.25 (0.96–1.64)	0.09
Heart failure					
Incidence rate	4.08	3.66	4.17	3.97	
HR (95% CI)					
Unadjusted	1.00 (reference)	0.92 (0.82–1.03)	1.01 (0.90–1.14)	0.98 (0.78–1.23)	0.91
Model 1*	1.00 (reference)	0.95 (0.85–1.06)	1.03 (0.91–1.16)	1.00 (0.80–1.25)	0.81
Model 2†	1.00 (reference)	0.98 (0.88–1.10)	1.00 (0.88–1.12)	1.00 (0.80–1.26)	0.95
Atrial fibrillation					
Incidence rate	6.83	6.09	7.67	7.06	
HR (95% CI)					
Unadjusted	1.00 (reference)	0.91 (0.84–1.00)	1.11 (1.01–1.22)	1.04 (0.88–1.25)	0.23
Model 1*	1.00 (reference)	0.93 (0.85–1.02)	1.12 (1.02–1.23)	1.07 (0.89–1.27)	0.17
Model 2†	1.00 (reference)	1.00 (0.92–1.09)	1.06 (0.96–1.16)	1.05 (0.88–1.25)	0.46

Incidence rates are n per 100 person-years or risk ratio (95% CI) unless noted otherwise. CCS indicates Canadian Cardiovascular Society; HR, hazard ratio.
*Model 1 includes adjustment for age, sex, and race/ethnicity.
†Model 2 includes variables in model 1 and additional adjustment for dyslipidemia, diabetes mellitus, hypertension, heart failure, smoking status, body mass index, and medication use at baseline (statins, β-blockers, long-acting nitrates, calcium channel blockers, aspirin, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers).

Costs

The mean total healthcare costs in the 1-year follow-up period were $28 674, $29 347, $35 416, and $34 771 among those with CCS classes I, II, III and IV, respectively (Table 4). The fully adjusted, total incremental costs associated with classes II, III, and IV, relative to class I, were $1184 (95% CI, −$876 to $3244), $6160 (95% CI, $3720–8600), and $6006 (95% CI, $1239–10 773), respectively (*P* for trend=0.0009). Mean outpatient healthcare costs among CCS classes I, II, III, and IV were $12 934, $12 856, $14 387, and $14 257, respectively. Mean inpatient healthcare costs among CCS classes I, II, III, and IV were $13 010, $13 932, $18 113, and $17 842, respectively. Fully adjusted pharmacy costs were similar between CCS class groups (*P*=0.98 for trend).



No. at Risk

CCS Class I	3983	3343	2761	2176	1604	1169	784	420
CCS Class II	5576	4818	4002	3279	2550	1940	1329	779
CCS Class III	3978	3248	2612	2049	1538	1093	784	461
CCS Class IV	679	554	454	372	275	198	144	92

Figure 4. Cumulative hazard of all-cause mortality by CCS class. The median follow-up time is 3.4 years (interquartile range: 1.6–5.6 years). CCS indicates Canadian Cardiovascular Society (angina classification).

Discussion

In this retrospective cohort analysis of veterans with stable angina, documented CCS classification in the electronic health record identified by NLP was positively associated with all-cause mortality and healthcare utilization for cardiovascular causes. Compared with CCS class I, those with CCS class III or IV angina symptom severity near the time of angina diagnosis experienced 33% to 48% higher all-cause mortality rate over median 3-year follow-up. All-cause mortality rates were similar among those presenting with class I and II symptoms. A graded association was also noted in which CCS class was associated with progressively higher risk of undergoing coronary revascularization (ie, PCI and coronary artery bypass grafting procedures). CCS classes III and IV were also significantly associated with more incurred direct healthcare costs. To our knowledge, the current analysis represents the first large, electronic health record–based cohort study reporting CCS classification and its association with all-cause mortality in newly diagnosed stable angina patients. These results support the prognostic importance of assessing and documenting angina severity at the point of care.

To date, several studies have shown mixed results regarding the association between angina severity and clinical outcomes, but these studies have generally been limited to hospitalized patients undergoing revascularization.[5,7,16–18] Clinical and survey data from ACQUIP (Veterans Affairs Medical Center Ambulatory Care Quality Improvement Project) demonstrated that greater angina severity scores, as measured by the Seattle Angina Questionnaire (SAQ), were associated with all-cause mortality and increased admission rates for acute coronary syndrome.[17,18] Conversely, a post hoc analysis from the BARI 2D (Bypass Angioplasty Revascularization Investigation 2 Diabetes) registry found similar rates of all-cause death and myocardial infarction among patients with CCS classes III and IV compared with those with CCS classes I and II.[19] Nonetheless, in light of the lack of consensus regarding the most appropriate method for assessing and documenting anginal symptoms at the point of care (ie, physician assessment with the CCS or patient-reported using the SAQ), objective evidence of myocardial ischemia (eg, inducible ischemia using exercise treadmill testing with stress echocardiography), as opposed to angina symptoms alone, is a stronger marker of adverse cardiovascular risk.[19,20] Although our study did not investigate CCS classification with objective evidence of myocardial ischemia on noninvasive testing, our findings justify the utility of CCS classification and documentation in routine practice. The relative merits, utility, and value of provider-documented symptom severity (eg, the CCS) compared with patient-reported symptom assessments (eg, the SAQ) remain to be fully understood.[21] Programs collecting patient-reported outcomes as part of routine clinical care, such as the University of Utah's mEVAL initiative, may provide insights into the relative value of provider- versus patient-reported outcome measures in contemporary practice.[22]

Our study found that veterans with higher CCS class were more likely to undergo revascularization procedures after CCS documentation; unexpectedly, however, we did not observe more aggressive medication therapy in those with more severe angina symptoms at baseline. The COURAGE (Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation) trial did not demonstrate a mortality benefit with revascularization via PCI compared with medical therapy alone for patients with stable angina.[23] Similarly, ORBITA (Objective Randomised Blinded Investigation with optimal medical Therapy of Angioplasty in stable angina) demonstrated no statistically significant difference in exercise time between patients with severe (≥70%) single-vessel coronary artery stenosis who underwent a PCI procedure compared with a sham procedure.[24] In context of the COURAGE and ORBITA results, our findings highlight an area for improving the extent to which medical therapy is optimized before attempting revascularization, especially in light of our finding of higher healthcare utilization and costs among those with more severe anginal symptoms (ie, CCS III and IV patients); however, additional research is needed on best practice strategies to ensure medical optimization before PCI.

Our analyses demonstrate that inpatient and healthcare costs are higher among veterans who present with more severe anginal symptoms. In contrast to other studies assessing angina-associated healthcare costs, the greatest proportion of total healthcare costs observed were incurred during inpatient care, followed by outpatient and pharmacy costs.[25] Notably, pharmacy costs did not differ significantly across CCS classes; this result may reflect greater use of

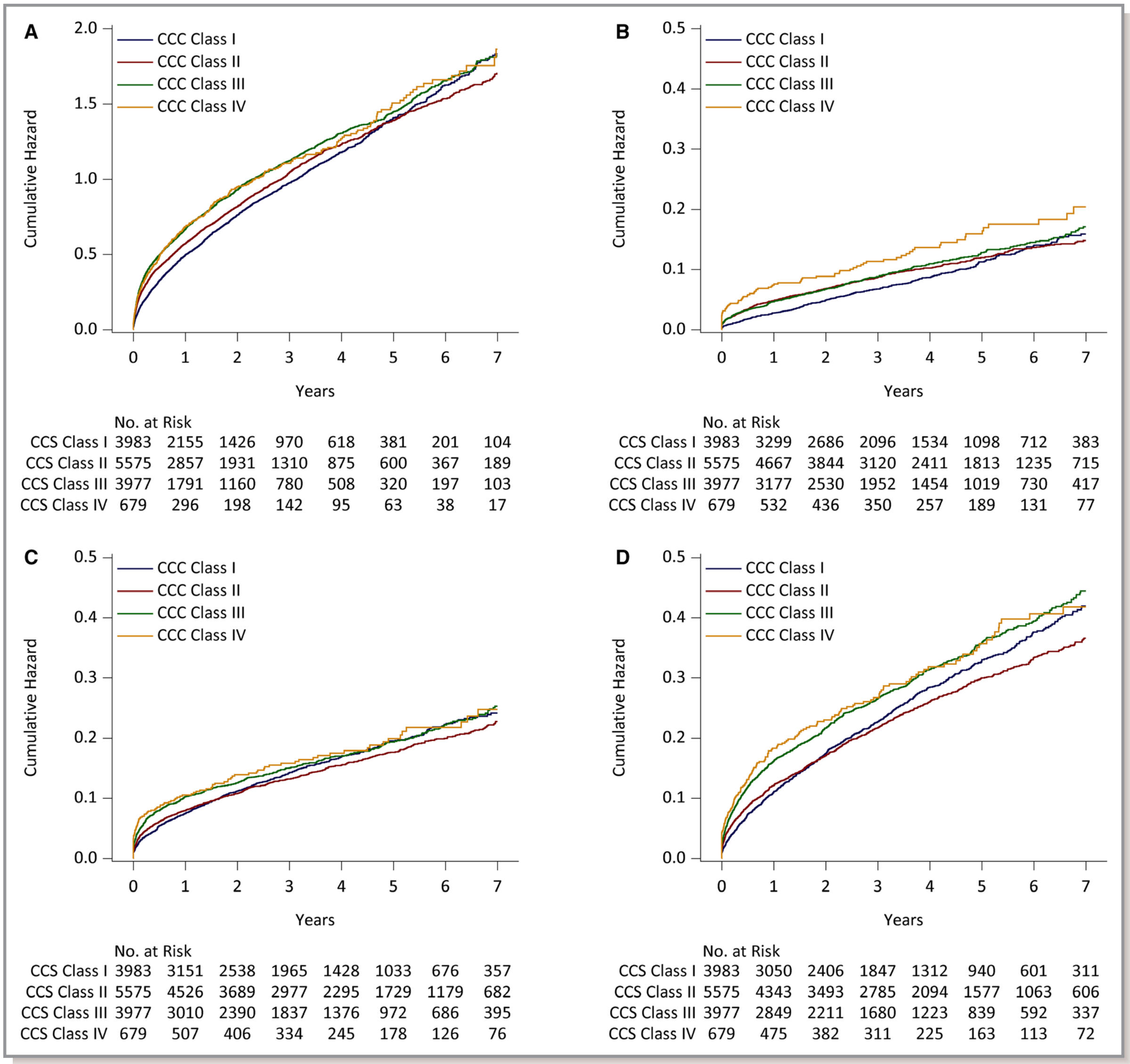


Figure 5. Cumulative hazard of all-cause and cause-specific hospitalizations by CCS class. **A**, All-cause hospitalizations. **B**, Hospitalizations due to acute coronary syndrome. **C**, Hospitalizations due to atrial fibrillation. **D**, Hospitalizations due to heart failure. CCS indicates Canadian Cardiovascular Society (angina classification).

revascularization strategies over medical therapy in patients with more debilitating angina.

Our study has several strengths, including the large sample size and use of NLP for extraction of CCS classifications from free-text medical notes from a large integrated healthcare database. The use of NLP in research settings to identify exposures is increasing and has been used in recent years to identify critical limb ischemia and bleeding based on clinical notes,[26,27] to predict mortality based on intensive care nursing notes,[28] to recognize pneumonia among chest radiograph reports,[29] to estimate medication dosing,[30–32] and more. The VHA is the largest integrated delivery network in the United States and thus offers the advantage of capturing comprehensive inpatient and outpatient healthcare encounters, laboratory, procedural, billing, and pharmacy data. Our study has several limitations worth noting. The use of CCS classification itself has a number of weaknesses because it is ultimately a physician-based assessment rather than a patient-reported outcome, such as the SAQ.[19,33] In addition, in our study, only a small proportion of the entire incident angina population had a CCS classification recorded (≈5%) despite the high precision of the NLP tool (93%). This finding introduces the possibility of an inherent bias in the documentation of CCS class toward sicker patients who may be more likely to undergo revascularization procedures. The recall of our tool, although high at 75%, also introduces potential bias because ≈25% of documented CCS cases could not be identified with our NLP tool and thus could not be included in the analysis. When comparing patients with and without CCS classification available, a larger percentage of

Table 3. Incidence Rates and HRs for Revascularization Procedures by CCS Classification

Outcomes	CCS Classification I (n=3983)	II (n=5576)	III (n=3978)	IV (n=679)	*P* Value For Trend
PCI					
Incidence rate	1.13	1.64	2.12	2.22	
HR (95% CI)					
Unadjusted	1.00 (reference)	1.50 (1.24–1.81)	1.83 (1.51–2.23)	1.94 (1.41–2.67)	<0.0001
Model 1*	1.00 (reference)	1.48 (1.22–1.79)	1.82 (1.50–2.21)	1.91 (1.39–2.63)	<0.0001
Model 2†	1.00 (reference)	1.51 (1.25–1.83)	1.87 (1.53–2.27)	1.92 (1.40–2.64)	<0.0001
CABG					
Incidence rate	1.87	3.20	4.40	4.93	
HR (95% CI)					
Unadjusted	1.00 (reference)	1.75 (1.51–2.03)	2.20 (1.90–2.55)	2.52 (2.01–3.18)	<0.0001
Model 1*	1.00 (reference)	1.73 (1.49–2.00)	2.20 (1.89–2.55)	2.49 (1.98–3.14)	<0.0001
Model 2†	1.00 (reference)	1.72 (1.48–1.99)	2.26 (1.94–2.62)	2.51 (1.99–3.16)	<0.0001

Incidence rates are n per 100 person-years or risk ratio (95% CI) unless noted otherwise. CABG indicates coronary artery bypass grafting; CCS, Canadian Cardiovascular Society; HR, hazard ratio; PCI, percutaneous coronary intervention.
*Model 1 includes adjustment for age, sex, and race/ethnicity.
†Model 2 includes variables in model 1 and additional adjustment for dyslipidemia, diabetes mellitus, hypertension, heart failure, smoking status, body mass index, and medication use at baseline (statins, β-blockers, long-acting nitrates, calcium channel blockers, aspirin, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers).

patients without CCS classification had documentation by a cardiologist. In addition, common comorbidities (eg, diabetes mellitus and hypertension) were more prevalent in those with a CCS classification than in those without, and Charlson comorbidity index was higher (indicating more comorbidity burden) in those without a CCS classification. Consequently, those veterans included in our study may not be representative of the larger angina population. Furthermore, the external validity of our findings is limited by the use of a single healthcare system with a nearly exclusively male and largely white population. Our study cannot address the important differences in angina symptoms based on race or sex, which have been described.[17] Given that we used VHA administrative data for utilization outcomes, we did not capture utilization that occurred outside the VHA. Finally, as with all observational studies, although we adjusted for important confounders when assessing outcomes, unmeasured variables may affect our results.

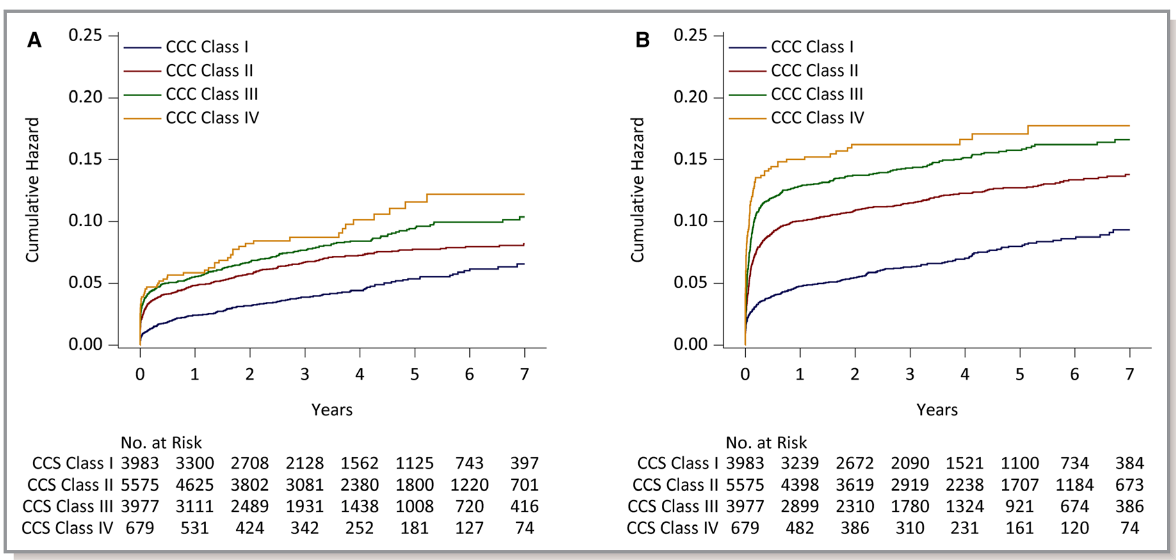


Figure 6. Cumulative hazard of revascularizations. **A**, Percutaneous coronary intervention (PCI) procedures. **B**, Coronary artery bypass grafting (CABG) procedures. CCS indicates Canadian Cardiovascular Society (angina classification).

Table 4. One-Year Total and Incremental Healthcare Costs by CCS Classification

Costs*	CCS Classification				*P* Value for Trend
	I (n=3983)	II (n=5576)	III (n=3978)	IV (n=679)	
Total (all settings)					
Cost for 1 y, mean (median)	28 674 (13 187)	29 347 (14 229)	35 416 (17 448)	34 771 (17 771)	
Incremental cost (95% CI)					
Unadjusted	1.00 (reference)	663 (−1653 to 2978)	6759 (3973–9544)	6111 (675–11 548)	0.002
Model 1*	1.00 (reference)	671 (−1600 to 2942)	6443 (3724–9161)	6386 (1020–11 751)	0.001
Model 2†	1.00 (reference)	1184 (−876 to 3244)	6160 (3720–8600)	6006 (1239–10 773)	0.0009
Outpatient					
Cost for 1 y, mean (median)	12 934 (8302)	12 856 (8411)	14 387 (9424)	14 257 (9579)	
Incremental cost (95% CI)					
Unadjusted	1.00 (reference)	−83 (−679 to 513)	1458 (775–2141)	1328 (27–2629)	0.005
Model 1*	1.00 (reference)	−129 (−725 to 467)	1390 (708–2072)	1323 (21–2625)	0.005
Model 2†	1.00 (reference)	168 (−422 to 757)	1426 (756–2096)	1319 (49–2590)	0.006
Inpatient					
Cost for 1 y, mean (median)	13 010 (0)	13 932 (0)	18 113 (0)	17 842 (0)	
Incremental cost (95% CI)					
Unadjusted	1.00 (reference)	918 (−1130 to 2965)	5114 (2510–7718)	4842 (−427 to 10 110)	0.01
Model 1*	1.00 (reference)	1095 (−831 to 3022)	4962 (2534–7389)	5379 (308–10 450)	0.004
Model 2†	1.00 (reference)	1299 (−468 to 3066)	4742 (2544–6940)	5050 (531–9570)	0.003
Pharmacy					
Cost for 1 y, mean (median)	2730 (1661)	2559 (1635)	2916 (1894)	2671 (1867)	
Incremental cost (95% CI)					
Unadjusted	1.00 (reference)	−172 (−363 to 19)	187 (−32 to 406)	−58 (−443 to 328)	0.76
Model 1*	1.00 (reference)	−164 (−347 to 19)	179 (−31 to 389)	−62 (−430 to 307)	0.79
Model 2†	1.00 (reference)	−174 (−365 to 17)	91 (−124 to 307)	−83 (−463 to 297)	0.98

All costs are in 2013 dollars. CCS indicates Canadian Cardiovascular Society.
*Model 1 includes adjustment for age, sex, and race/ethnicity.
†Model 2 includes variables in model 1 and additional adjustment for dyslipidemia, diabetes mellitus, hypertension, heart failure, smoking status, body mass index, and medication use at baseline (statins, β-blockers, long-acting nitrates, calcium channel blockers, aspirin, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers).

Conclusion

Our study found that greater angina severity at the time of angina diagnosis, as measured by CCS classification identified from clinical notes using NLP techniques, was associated with higher rates of mortality, cardiovascular hospitalizations, coronary revascularization, and overall healthcare costs. Furthermore, these outcomes were largely graded, with stronger associations present as severity increased, and persisted after controlling for variables known to affect differences in angina symptoms, prognosis, and treatment.[17] These findings are of particular relevance to clinicians' efforts to judiciously utilize revascularization interventions, to improve patients' symptomatic burden from angina, and to reduce overall healthcare costs. The results of the current analysis support the prognostic importance of assessing and documenting angina severity at the point of care.

Acknowledgments

We acknowledge Patrick Alba for his help in the development of the natural language processing system and data analysis.

Sources of Funding

This work was supported using resources and facilities at the Veterans Affairs (VA) Salt Lake City Health Care System with funding from VA Informatics and Computing Infrastructure VA HSR HIR 08-204 and the University of Utah and by an investigator-initiated grant from Gilead Sciences.

Disclosures

Drs Bress, LaFleur, and Nelson and Mr Crook received support from the Investigator Sponsored Research Grant from Gilead Sciences. Drs Dodson and King currently serve as consultants

for Novartis Pharmaceuticals on a research project unrelated to the current study. Dr Dodson is supported by a Patient Oriented Career Development Award (K23AG052463) from the National Institute of Health/National Institute of Aging and a Mentored Clinical and Population Research Award from the American Heart Association. Dr Shah is supported by a grant from the National Heart, Lung, and Blood Institute (K08HL136850). The remaining authors have no disclosures to report.

References

1. Mensah GA, Brown DW. An overview of cardiovascular disease burden in the United States. *Health Aff (Millwood)*. 2007;26:38–48.
2. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB III, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR, Smith SC, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: executive summary A Report of the American College of Cardiology Foundation/ American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *Circulation*. 2012;126:3097–3137.
3. Kannel WB, Feinleib M. Natural history of angina pectoris in the Framingham study: prognosis and survival. *Am J Cardiol*. 1972;29:154–163.
4. Campeau L. Letter: grading of angina pectoris. *Circulation*. 1976;54:522–523.
5. Kaul P, Naylor CD, Armstrong PW, Mark DB, Theroux P, Dagenais GR. Assessment of activity status and survival according to the Canadian Cardiovascular Society angina classification. *Can J Cardiol*. 2009;25:e225–e231.
6. Weinstein G, Goldbourt U, Tanne D. Angina pectoris severity among coronary heart disease patients is associated with subsequent cognitive impairment. *Alzheimer Dis Assoc Disord*. 2015;29:6–11.
7. Hemingway H, Fitzpatrick N, Gnani S, Feder G, Walker N, Crook A, Magee P, Timmis A. Prospective validity of measuring angina severity with Canadian Cardiovascular Society class: the ACRE study. *Can J Cardiol*. 2004;20:305–309.
8. Phelps CE, Buysman EK, Gomez Rey G. Costs and clinical outcomes associated with use of ranolazine for treatment of angina. *Clin Ther*. 2012;34:1395–1407.e4.
9. Kent ST, Safford MM, Zhao H, Levitan EB, Curtis JR, Kilpatrick RD, Kilgore ML, Muntner P. Optimal use of available claims to identify a Medicare population free of coronary heart disease. *Am J Epidemiol*. 2015;182:808–819.
10. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. *Nat Lang Eng*. 2004;10:327–348.
11. Cornia R, Patterson OV, Ginter T, DuVall SL. Rapid NLP Development with Leo. 2014.
12. DuVall SL, Cornia R, Forbush TB, Halls CH, Patterson OV. Check it with Chex: A Validation Tool for Iterative NLP Development. In: AMIA Symposium Proceedings. *AMIA*; 2014.
13. South BR, Shen S, Leng J, Forbush TB, Duvall SL, Chapman WW. A Prototype Tool Set to Support Machine-Assisted Annotation. In: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing (BioNLP 2012). Association for Computational Linguistics; 2012:130–139.
14. Leng J, Shen S, Gundlapalli A, South B. eHOST: The Extensible Human Oracle Suite of Tools for Annotation of Clinical Narratives [abstract]. Phoenix, AZ; 2010.
15. Quan H, Sundarajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–1139.
16. Steg PG, Greenlaw N, Tendera M, Tardif JC, Ferrari R, Al-Zaibag M, Dorian P, Hu D, Shalnova S, Sokn FJ, Ford I, Fox KM; Prospective Observational Longitudinal Registry of Patients with Stable Coronary Artery Disease Investigators. Prevalence of anginal symptoms and myocardial ischemia and their effect on clinical outcomes in outpatients with stable coronary artery disease: data from the international observational CLARIFY registry. *JAMA Intern Med*. 2014;174:1651–1659.
17. Mozaffarian D, Bryson CL, Spertus JA, McDonell MB, Fihn SD. Anginal symptoms consistently predict total mortality among outpatients with coronary artery disease. *Am Heart J*. 2003;146:1015–1022.
18. Spertus JA, Jones P, McDonell M, Fan V, Fihn SD. Health status predicts long-term outcome in outpatients with coronary disease. *Circulation*. 2002;106:43–49.
19. Dagenais GR, Lu J, Faxon DP, Bogaty P, Adler D, Fuentes F, Escobedo J, Krishnaswami A, Slater J, Frye RL; BARI 2D Study Group. Prognostic impact of the presence and absence of angina on mortality and cardiovascular outcomes in patients with type 2 diabetes and stable coronary artery disease: results from the BARI 2D (Bypass Angioplasty Revascularization Investigation 2 Diabetes) trial. *J Am Coll Cardiol*. 2013;61:702–711.
20. Gehi AK, Ali S, Na B, Schiller NB, Whooley MA. Inducible ischemia and the risk of recurrent cardiovascular events in outpatients with stable coronary heart disease: the Heart and Soul Study. *Arch Intern Med*. 2008;168:1423–1428.
21. Rumsfeld JS, Alexander KP, Goff DC, Graham MM, Ho PM, Masoudi FA, Moser DK, Roger VL, Slaughter MS, Smolderen KG, Spertus JA, Sullivan MD, Treat-Jacobson D, Zerwic JJ; American Heart Association Council on Quality of Care and Outcomes Research, Council on Cardiovascular and Stroke Nursing, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease, Stroke Council. Cardiovascular health: the importance of measuring patient-reported health status: a scientific statement from the American Heart Association. *Circulation*. 2013;127:2233–2249.
22. Biber J, Ose D, Reese J, Gardiner A, Facelli J, Spuhl J, Brodke D, Lee VS, Hess R, Weeks H. Patient reported outcomes—experiences with implementation in a University Health Care setting. *J Patient Rep Outcomes*. 2018;2:34.
23. Weintraub WS, Boden WE, Zhang Z, Kolm P, Zhang Z, Spertus JA, Hartigan P, Veledar E, Jurkovitz C, Bowen J, Maron DJ, O'Rourke R, Dada M, Teo KK, Goeree R, Barnett PG; Department of Veterans Affairs Cooperative Studies Program No. 424 (COURAGE Trial) Investigators and Study Coordinators. Cost-effectiveness of percutaneous coronary intervention in optimally treated stable coronary patients. *Circ Cardiovasc Qual Outcomes*. 2008;1:12–20.
24. Al-Lamee R, Thompson D, Dehbi H-M, Sen S, Tang K, Davies J, Keeble T, Mielewczik M, Kaprielian R, Malik IS, Nijjer S, Petraco R, Cook C, Ahmad Y, Howard J, Baker C, Sharp A, Gerber R, Talwar S, Assomull R, Mayet J, Wensel R, Collier D, Shun-Shin M, Thom SA, Davies JE, Francis DP. Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial. *Lancet*. 2018;391:31–40.
25. Javitz HS, Ward MM, Watson JB, Jaana M. Cost of illness of chronic angina. *Am J Manag Care*. 2004;10:S358–S369.
26. Afzal N, Mallipeddi VP, Sohn S, Liu H, Chaudhry R, Scott CG, Kullo IJ, Arruda-Olson AM. Natural language processing of clinical notes for identification of critical limb ischemia. *Int J Med Inform*. 2018;111:83–89.
27. Taggart M, Chapman WW, Steinberg BA, Ruckel S, Pregenzer-Wenzler A, Du Y, Ferraro J, Bucher BT, Lloyd-Jones DM, Rondina MT, Shah RU. Comparison of 2 natural language processing methods for identification of bleeding among critically ill patients. *JAMA Netw Open*. 2018;1:e183451.
28. Waudby-Smith IER, Tran N, Dubin JA, Lee J. Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. *PLoS One*. 2018;13:e0198687.
29. Jones BE, South BR, Shao Y, Lu CC, Leng J, Sauer BC, Gundlapalli AV, Samore MH, Zeng Q. Development and validation of a natural language processing tool to identify patients treated for pneumonia across VA emergency departments. *Appl Clin Inform*. 2018;9:122–128.
30. Nelson SD, Lu CC, Teng CC, Leng J, Cannon GW, He T, Zeng Q, Halwani A, Sauer B. The use of natural language processing of infusion notes to identify outpatient infusions. *Pharmacoepidemiol Drug Saf*. 2015;24:86–92.
31. Lu C-C, Leng J, Cannon GW, Zhou X, Egger M, South B, Burningham Z, Zeng Q, Sauer BC. The use of natural language processing on narrative medication schedules to computer average weekly dose. *Pharmacoepidemiol Drug Saf*. 2016;25:1414–1424.
32. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, Williams AE, Ross T, Wagner EH, Neslund-Dudas C, Altschuler A, Quinn V, Hornbrook M, Nekhlyudov L. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care*. 2009;47:1091–1097.
33. Cutrona SL, Toh S, Iyer A, Foy S, Daniel GW, Nair VP, Ng D, Butler MG, Boudreau D, Forrow S, Goldberg R, Gore J, McManus D, Racoosin JA, Gurwitz JH. Validation of acute myocardial infarction in the Food and Drug Administration's Mini-Sentinel program. *Pharmacoepidemiol Drug Saf*. 2013;22:40–54.

SUPPLEMENTAL MATERIAL

Data S1.

SUPPLEMENTAL METHODS

Angina NLP System Overview

To meet the goals of the project, a natural language processing (NLP) system was developed to identify instances of the Canadian Cardiovascular Society (CCS) angina score explicitly stated within clinical notes. All clinical documents available for a cohort of 6,556,919 VA patients were used to train and validate the system. A training set was created for a random set of patients within cohort. This set contained 1,885,840 documents and 7,456 patients. These documents were used to create the knowledge base for the NLP system.

NLP Pipeline Methods

The NLP system is based on the Unstructured Information Management Architecture Asynchronous Scaleout (UIMA AS; Ferrucci & Lally, 2004). The system also utilizes the VINCI Leo framework, and a set of libraries that enables more efficient utilization of UIMA AS (Cornia, et al, 2014).

The NLP pipeline was developed over several iterations that included cycles of system development and error analysis. The system design included several sets of manually created regular expressions in order to identify concepts related to CCS scores. The co-occurrence patterns of words and phrases that were found in text through the use of these regular expressions were used to disambiguate relevant concepts. The NLP system has the following modules:

1. Regular Expressions:

[CCS_Concept] – These expressions areused to identify mentions of CCS score such as “CCS”, “CCSC”, “Canadian”, “Functional class” and “angina”.

[Concept_word]– These expressions enable detection of words that are frequently used in a CCS phrase, such as “grade”, “class”, “score”, “stage”.

[CCS_Score] – These expressions capture numeric values typically associated with a score, such as 1,2,3,4,I,II,III,IV,1-2, 2-3, etc…)

[Exclude] – These expressions identify words and numbers that indicate that the surrounding phrase does not describe CCS score, such as dates, other irrelevant numeric values, and invalid units of measure.

2. Patterns:

After all of these various terms and characters are captured, two pattern types are used in order to separate relevant phrases from the irrelevant. These two patterns are CCS_Pattern and Exclude_Pattern.

[CCS_Pattern] – These patterns describe valid sequences of CCS_Concept, Concept_word, and CCS Score that can be encountered in the text. Examples of such sequences are outlined in Supplemental Table 4.

A few very specific patterns were created in order to capture special cases like questionnaires and form, such as “ [x] Angina 3?”

[Exclude_Pattern] – These patterns describe sequences of previously captured phrases that indicate that are actually irrelevant and do not actually represent CCS score. For example, a phase “Angina 2x/week” contains CCS_concept phrase "angina" and

CCS_score phrase "2", but this phrase is detected as irrelevant because it also contains an Exclude phrase "x/week", which indicates that the phrase is irrelevant. Other examples of phrases captured by exclusion patterns are “Angina 2/2” , “Angina 2/13/12”, and “1000ccs given to pt”.

SUPPLEMENTAL RESULTS

All clinical notes for all 6,556,919 patients in the cohort were processed. The final output contained 64,732 mentions of CCS scores for 22,994 patients. NLP systems are typically evaluated using precision, recall, and F-1 score performance metrics.

For precision validation, we selected 500 documents that contained CCSscore mentions identified by the NLP system. The validation set had 609 mentions that were manually reviewed. The validation step was completed using the Chex validation tool (DuVall, et al 2014.). This validation found that the NLP system had 93.1% precision (positive predictive value).

Due to the extreme rarity of the CCS concept mentions in clinical notes, an enriched set of notes was needed in order to keep human annotators from reading over 1000 documents before finding a single instance of a CCS score. Sampling was done using a set of filtering criteria. First, the documents were filtered by only using documents from the date of an angina diagnosis, determined by ICD-9 codes. Using this single filter, we found that CCS scores were found in less than 0.5% of all documents. The corpus was further filtered by documents containing “card” in the document title (cardiology, cardiologist, etc.). Lastly, the presence of the word “angina” or “CAD” (coronary artery disease) was required to be in the documents. We were able to estimate that CCS could be found in roughly 6% of the documents after all of these filters were applied.

Using the filtering criteria, 500 documents identified for manual annotation. Three annotators reviewed the documents using the eHost annotation tool. They found 33 instances of a CCS score, 25 of which were also found by the system. This sensitivity analysis found that the NLP system had a 75.7% recall (sensitivity).

In addition to explicit CCS scores, the annotators also marked any instance of angina being described in terms of exertion, which could be used to infer CCS score (i.e "angina after 2 flights of stairs", "angina after walking two blocks."). In the same 500 documents, 240 instances of exertional angina descriptions were found. These statements were not used for system development but can serve as a basis for future work.

Limitations and Future Work

The rarity of the CSS score mentions in clinical notes leads to several limitations:

First, even after being enriched with several filters, the recall corpus of 500 manually reviewed documents found only 33 instances of a CCS score. Such a low number of CCS score mentions increases likelihood that the true recall differs from the measured value.

Second, the set of filters used to select documents for review introduce a risk of selection bias.

Error analysis revealed that a relatively low recall was due to several causes. First, new abbreviations appeared in the selected documents that were not found in the training set. Angina Pectoris was abbreviated twice as AP (i.e. "AP class II") and the Functional class was abbreviated as FC (i.e. "pt complains of angina FC III"). Another cause for a lower than expected recall score was the instance of novel spellings that had not been encountered during the knowledge base acquisition. In one instance, “CCC class I” was found rather than the expected CCS.

Future work may include the following steps: 1) With additional effort, another development iteration can be performed to improve the recall and precision.
2) CCS score can be potentially inferred using exertional angina descriptions. The manual annotations performed for sensitivity analysis can be used as a training set for a potential CCS score inference algorithm.

Table S1. Phelp's Criteria for classification of chronic stable angina. The protocol uses three definitions (Criteria A, B or C) to identify angina veterans from a combination of ICD-9-CM diagnostic and procedural codes, CPT codes, and medication use.

Criteria	***Criteria requirements***
A	≥ 2 diagnosis codes for CAD on separate dates within a 12-month period
	≥ 2 diagnosis codes for chest pain on separate dates
	≥ 2 nitrate prescriptions with a gap of at least 30 days between fill dates
B	≥ 1 diagnosis code for stable angina
	≥ 2 prescription fills for nitrates with a gap of at least 30 days between fills within a 12-month period
C	≥ 2 diagnosis codes for stable angina on separate dates (one if it is a primary diagnosis during an inpatient stay)
	Did not have at least two pharmacy fills for nitrates with a gap of at least 30 days
	Either ≥ 2 beta-blocker prescriptions or ≥ 2 calcium channel blocker prescriptions with a gap of at least 30 days between fills within a 12-month period

Condition	Code type	Code/Description
Chest pain	ICD-9-CM diagnosis	786.5 (Chest pain, unspecified), 786.51 (Precordial pain), 786.52 (Painful respiration), 786.59 (Other chest pain)

Stable angina	ICD-9-CM diagnosis	413.x (Angina pectoris)
CAD	ICD-9-CM diagnosis	410.xx (Acute myocardial infarction), 411.xx (Other acute and sub-acute forms of ischemic heart disease), 412 (Old myocardial infarction), 414.xx (Other forms of chronic ischemic heart disease), 429.2 (Cardiovascular disease, unspecified), 429.5 (Rupture of chordae tendinae), 429.6 (Rupture of papillary muscle), 429.7x (Certain sequelae of myocardial infarction, not elsewhere specified), 996.03 (Mechanical complications due to coronary bypass graft), V45.81 (Aortocoronary bypass status), V45.52 (Percutaneous transluminal coronary angioplasty status)
	ICD-9-CM procedure	0.66 (Percutaneous transluminal coronary angioplasty or atherectomy), 36.0x (Removal of coronary artery obstruction and insertion of stent), 36.1x (Bypass anastomosis for heart revascularization), 36.2 (Heart revascularization by arterial implant), 36.3x (Other heart revascularization)

	CPT	33140-33141 (Transmyocardial laser revascularization, by thoracotomy), 33510-33523, 33533-33536 (Coronary artery bypass), 33572 (Coronary endarterectomy in conjunction with CABG), 92975 (Thrombolysis, coronary; by intracoronary infusion), 92980-92981 (Transcatheter placement of an intracoronary stent (s), percutaneous, with or without other therapeutic intervention, any method), 92982, 92984 (Percutaneous transluminal coronary balloon angioplasty), 92995-92996 (Percutaneous transluminal coronary atherectomy, by mechanical or other method, with or without balloon angioplasty), 93540 (Injection procedure during cardiac catheterization for selective opacification of aortocoronary venous bypass grafts)
	HCPCS	G0290-G0291 (Transcatheter placement of drug eluting intracoronary stent), S0340-S0342 (Lifestyle modification program for management of CAD), S2205-S2209 (Minimally invasive direct coronary artery bypass surgery involving mini-thoracotomy or mini-sternotomy surgery using arterial and/or venous graft), G8033-G8041, G8498 (Patient coded as a coronary artery disease patient), G8159-G8172, G9497 (Patient coded as a coronary artery bypass graft patient)

CABG = Coronary artery bypass grafting; CAD = Coronary artery disease; CPT= Current Procedural Terminology; HCPCS = Healthcare Common Procedure Coding System; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification

*Adapted with permission from Phelps CE, Buysman EK, Gomez Rey G. Costs and clinical outcomes associated with use of ranolazine for treatment of angina. *Clin Ther*. 2012;34:1395–1407.e4.

Table S2. Examples of used patterns and phrases that match the patterns

Pattern	Example
[CCS_word] [CCS_score] [CCS_Concept]	“class IV angina” “grade 2 CCS”
[CC_Concept] [CCS_Word] [CCS_score]	“Angina class 2” “Canadian score 3”
[CCS_concept] [CCS_score]	"CCS 1" "functional class 3" "angina IV"

Table S3. Codes used to define variables in the analysis

Variable	Code(s)
Angina	≥1 claim with an ICD-9-CM diagnosis (any position) of 413.x.
Chest pain	≥1 claim with an ICD-9-CM diagnosis (any position) of 786.50, 786.51, 786.52, and 786.59.
Coronary artery disease	*Any one of the following:* a) ≥1 claim with an ICD-9-CM diagnosis (any position) of 410.xx, 411.x, 412, 414.xx, 429.2, 429.5, 429.6, 429.7x, 996.03, V45.81, V45.82 b) ≥1 claim with an ICD-9-CM procedure (any position) of 00.66, 36.0x, 36.1x, 36.2, 36.3x c) ≥1 claim with a CPT code (any position) of 33140, 33141, 33510-33523, 33533-33536, 33572, 92975, 9280, 92981, 92982, 92984,92995, 92996, 93540 d) ≥1 claim with a HCPCS code (any position) of G0290, G0291, S0340-S0342, S2205-S2209, G8033-G8041, G8498, G8159-G8172, G9497
Baseline characteristics	

Dyslipidemia	≥1 claim with an ICD-9-CM diagnosis (any position) of 272.4
Diabetes mellitus	≥1 claim with an ICD-9-CM diagnosis (any position) of 249.xx, 250.xx, 357.2, 362.0x, 366.41
Myocardial infarction	*Any one of the following:* a) Acute myocardial infarction: ≥1 claim with an ICD-9-CM diagnosis (any position) of 410.xx, b) History of myocardial infarction: ≥1 claim with an ICD-9-CM diagnosis (any position) of 411.0, 412
Hypertension	≥1 claim with an ICD-9-CM diagnosis (any position) of 401.xx, 402.xx, 403.xx, 404.xx, 405.xx
Heart failure	≥1 claim with an ICD-9-CM diagnosis (any position) of 398.91, 402.01, 402.11, 402.91 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.xx
Smoking	≥1 claim with an ICD-9-CM diagnosis (any position) of 305.1, 649.0x, 989.84, V15.82
Atrial fibrillation	≥1 claim with an ICD-9-CM diagnosis (any position) of 427.31

6-month pre-index costs	Sum of the inpatient and outpatient total costs. Use the variable acttotcost from DSS Outpatient for outpatient costs and the variable totcostin from the DSS Discharge dataset for inpatient costs.
6-month pre-index ED visits	Identified by primary or secondary clinic stop codes 101 (prior to 2007) and 130 and 131 (since 2007). Multiple ED codes on the same service date counted as one visit.
6-month pre-index hospitalizations	Numerical variable of the number of hospitalizations by the patient in the 6-month pre-index period.
Healthcare utilization metrics	
Total costs	Sum of the inpatient and outpatient total costs. Use the variable acttotcost from DSS Outpatient for outpatient costs and the variable totcostin from the DSS Discharge dataset for inpatient costs.
Outpatient costs	Sum of the outpatient total costs using the variable acttotcost from DSS Outpatient.
Inpatient costs	Sum of the inpatient total costs using the variable totcostin from DSS Discharge dataset.
Pharmacy costs	Sum of the total pharmacy costs using DSS.

Percutaneous coronary intervention	*Any one of the following:* a) ≥1 claim with a CPT code (any position) of 92980, 92981, 92982, 92984, 92995, 92996, 92920, 92921, 92924, 92925, 92928, 92929, 92933, 92934, 92937, 92938, 92941, 92943, 92944, 92973, 92973, 92975, 92978, 92979. b) ≥1 claim with a HCPCS code (any position) of G0290-G0291, C9600-C9608 c) ≥1 claim with an ICD-9-CM procedure code (any position) of 00.66, 36.01, 36.02, 36.05, 36.06, 36.07
Coronary artery bypass graft	*Any one of the following:* a) ≥1 claim with a CPT code (any position) of 33510-33516, 33517-33523, 33533-33536 b) ≥1 claim with an HCPCS code (any position) of S2205-S2209 c) ≥1 claim with an ICD-9-CM procedure code (any position) of 36.1x.
Emergent percutaneous coronary intervention	A percutaneous coronary intervention with a corresponding hospitalization and ICD-9-CM of 410.x. This method has been previously been validated. (30)

All-cause hospitalizations	Identified as an inpatient visit (Medical or Surgical) in the CDW.
Acute coronary syndrome hospitalizations	≥1 claim with an ICD-9-CM diagnosis (any position) of 410.00-410.92 associated with an inpatient visit
Heart failure hospitalizations	≥1 claim with an ICD-9-CM diagnosis (any position) of 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0-428.9 associated with an inpatient visit
Atrial fibrillation hospitalizations	≥1 claim with an ICD-9-CM diagnosis (any position) of 427.31 associated with an inpatient visit
Emergency department visits	Post-index period ED visits were identified by primary or secondary clinic stop codes 101 (prior to 2007) and 130 and 131 (since 2007) in the CDW and counted. Multiple ED codes on the same service date were counted as one visit.
Outpatient visits	Outpatient visits were identified by primary or secondary clinic stop codes in the CDW and included all outpatient visits including physician visits, laboratory, radiology, physical therapy, etc.

General medicine outpatient visits	Outpatient visits were identified by primary or secondary clinic stop codes 301, 323, 348, or 27 in the CDW and included all outpatient visits including physician visits, laboratory, radiology, physical therapy, etc.
Cardiology outpatient visits	Cardiology Outpatient visits were identified by primary or secondary clinic stop codes 303 in the CDW and included all outpatient visits including physician visits, laboratory, radiology, physical therapy, etc.

CDW = Corporate Data Warehouse; CPT = Current Procedural Terminology; DSS = Decision Support System; HCPCS = Healthcare Common Procedure Coding System; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification.

Table S4. Characteristics of Angina Patients With and Without a CCS Class Documentation

	CCS Classification		
Characteristic	**Without (n=276,600)**	**With (n=14,216)**	**p Value**
Age, y	67 ± 11	67 ± 9.8	<.001
Male Sex	268877 (97%)	13996 (98%)	<.001
Race			
White	214867 (78%)	11464 (81%)	<.001
Black	35765 (13%)	1662 (12%)	<.001
Asian	3721 (1%)	240 (2%)	<.001
American Indian	2542 (1%)	135 (1%)	0.705
Unknown/Missing	19705 (7%)	715 (5%)	<.001
Census Region			
Northeast	31469	1387	<.001
Midwest	65717	2240	<.001
South	125785	6948	<.001
West	46683	2365	0.454
Missing/Other	6946	1276	<.001
Pre-index comorbidities and clinical characteristics			
CCI Score	1.3 ± 2.2	1.1 ± 2.0	<.001
Dyslipidemia	153955 (56%)	8670 (61%)	<.001

Diabetes Mellitus	114191 (41%)	6650 (47%)	<.001
Acute or history of MI	21138 (8%)	0 (0%)	
Hypertension	204123 (74%)	11380 (80%)	<.001
Heart Failure	51430 (19%)	2677 (19%)	0.476
Smoking	60761 (22%)	2949 (21%)	<.001
Alcohol Abuse	19407 (7%)	838 (6%)	<.001
Body Mass Index	30 ± 6.3	31 ± 5.9	<.001
BMI Categories			
Underweight	2901 (1%)	98 (1%)	<.001
Normal	49878 (18%)	2158 (15%)	<.001
Overweight	95688 (35%)	5083 (36%)	0.006
Obese	127875 (46%)	6877 (48%)	<.001
Medications			
Number of Baseline Anti-Anginals	1.3 ± 0.80	1.3 ± 0.86	<.001
Beta-blocker	196028 (71%)	10233 (72%)	0.005
Calcium channel blocker	62905 (23%)	3467 (24%)	<.001
Long Acting Nitrate	96800 (35%)	5354 (38%)	<.001
Aspirin	95882 (35%)	4929 (35%)	
Statin	198773 (72%)	10196 (72%)	0.715
ACE Inhibitor or ARB	164056 (59%)	8665 (61%)	<.001
P2Y12 Antagonist	67037 (24%)	3226 (23%)	<.001

Yes	110392 (40%)	7438 (52%)	<.001
No	122813 (44%)	4462 (31%)	<.001
Unknown/Missing	43395 (16%)	2316 (16%)	0.051
Seeing a cardiologist	173269 (63%)	7975 (56%)	<.001
Preindex Resource Utilization and Costs			
Hospitalizations			
Patients having	85243 (31%)	3552 (25%)	<.001
Number of Hospitalizations	0.54 ± 1.1	0.37 ± 0.78	<.001
ER events			
Patients Having	107567 (39%)	8202 (58%)	<.001
Number of Events	0.96 ± 1.9	3.7 ± 7.2	<.001
Total Costs	16221 ± 40274	35148 ± 95986	<.001

Data are number of patients (%) or mean (SD). CCI: Charlson/Quan Comorbidity Index, PCI=percutaneous coronary intervention, CABG=coronary artery bypass graft, CCB=calcium channel blocker, ACE=Angiotensin converting enzyme, ARB= Angiotensin II Receptor Blocker

Supplemental References:

Cornia R, Patterson O V, Ginter T, Duvall SL. Rapid NLP Development with Leo. In: AMIA Annu Symp Proc.; 2014.

Duvall SL, Cornia RC, Forbush TB, Halls CH, Patterson O V. Check it with Chex : A Validation Tool for Iterative NLP Development. In: AMIA Annu Symp Proc.; 2014.

Ferrucci, D., & Lally, A. (2004). UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering, 10(3-4).

South, B. R., Leng, J., Anderson, K., Shen, S., Thibault, J., & DuVall, S. (2010). The Extensible Human Oracle Suite of Tools (eHOST) for Pre-Annotation of Clinical Narratives. In BioCreative: Critical Assessment of Information Extraction in Biology Annual Conference.