SUPPLEMENTAL MATERIAL

Table S1. Validated algorithm to ascertain obesity from electronic health record

Eligibility: Inclusion Criteria	**Adult patients (18 and older) with** **1) Height and weight measurement at a frequency based upon patient age from structured data field in physical examination section of electronic health record** **2) Assessment of body mass index (BMI = Weight / [Height]^2) to categorize individuals as follows** **1. BMI ≥ 30** **2. Most recent BMI ≥29 w/ any BMI ≥30 in the past year** **3. No BMI data available** **3) If no BMI data available, obesity define based upon** **1. ICD-9/10 diagnosis code or** **2. Problem list term**
Eligibility: Exclusion Criteria	**1) Most recent BMI < 30** **2) Most recent BMI ≤ 29 with no BMI ≥ 30 in the past year**
Frequency of weight/height measurements	**1) Weight measurement at a frequency based upon age** **a) Patients age ≥65, most recent weight in past 2 years** **b) Patients age 40-64, most recent weight in past 3 years** **c) Patients age 23-39, most recent weight in past 5 years** **d) Patients aged 22, most recent weight in past 4 years** **e) Patients aged 21, most recent weight in past 3 years** **f) Patients aged 18-20, most recent weight in past 2 years** **2) Height measurement at a frequency based upon age** **a) Patients aged ≥28, most recent height in past 10 years** **b) Patients aged 27, most recent height in past 9 years** **c) Patients aged 26, most recent height in past 8 years**

	d) **Patients aged 25, most recent height in past 7 years** e) **Patients aged 24, most recent height in past 6 years** f) **Patients aged 23, most recent height in past 5 years** g) **Patients aged 22, most recent height in past 4 years** h) **Patients aged 21, most recent height in past 3 years** i) **Patients aged 18-20, most recent height in past 2 years**
Data Sources	1) **Problem list terms (any prior)** 2) **ICD 9/10 diagnosis codes in prior 3 years** a) **Hospitalization or any outpatient visit** b) **Any primary or secondary codes** 3) **Physical exam / flow sheets**
Problem List Terms	**Obese; Obesity; Morbid Obesity; Simple Obesity; Body Mass Index 30 + -Obesity**
ICD-9 codes	**278.00: Obesity, unspecified** **278.01: Morbid obesity** **278.03: Obesity Hypoventilation syndrome**
ICD-10 Codes	**E66.01: Morbid (severe) obesity due to excess calories** **E66.09: Other obesity due to excess calories** **E66.1: Drug-induced obesity** **E66.2: Morbid (severe) obesity with alveolar hypoventilation** **E66.3: Overweight** **E66.8: Other obesity** **E66.9: Obesity, unspecified** **Z68.30: Body mass index (BMI) 30.0-30.9, adult** **Z68.31: Body mass index (BMI) 31.0-31.9, adult** **Z68.32: Body mass index (BMI) 32.0-32.9, adult** **Z68.33: Body mass index (BMI) 33.0-33.9, adult** **Z68.34: Body mass index (BMI) 34.0-34.9, adult** **Z68.35: Body mass index (BMI) 35.0-35.9, adult**

	Z68.36: Body mass index (BMI) 35.0-35.9, adult **Z68.37: Body mass index (BMI) 37.0-37.9, adult** **Z68.38: Body mass index (BMI) 38.0-38.9, adult** **Z68.39: Body mass index (BMI) 39.0-39.9, adult** **Z68.41: Body mass index (BMI) 40.0-44.9, adult** **Z68.42: Body mass index (BMI) 45.0-49.9, adult** **Z68.43: Body mass index (BMI) 50-59.9 , adult** **Z68.44: Body mass index (BMI) 60.0-69.9, adult** **Z68.45: Body mass index (BMI) 70 or greater, adult**
Cut off values for height and weight	**1) Max height: 84 inches** **2) Min height: 48 inches** **3) Max weight: 515 pounds** **4) Min weight: 60 pounds**
Chart Review Selection Criteria	**1) Blinded list of 630 patients (25 positive and 10 negative patients per primary care practice in the network)** **2) Review performed by a Research Nurse**
Chart Review Results	**Sensitivity: 98%** **Specificity: 97%** **Positive predictive value: 97%** **Negative predictive value: 96%**

BMI: body mass index

Table S2. ICD-9/10 codes used to define codified predictors not ascertained using validated algorithms

Variable	ICD-9/10 Codes
Myocardial Infarction	410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 412, 429.79, I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I21.9 , I21.A1, I21.A9, I22.0, I22.1, I22.2, I22.8, I22.9, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.7, I23.8, I24.1, I25.2,
Chronic Kidney Disease	250.4, 250.41, 250.42, 250.43, 403, 403.01, 403.1, 403.11, 403.9, 403.91, 404, 404.01, 404.02, 404.03, 404.1, 404.11, 404.12, 404.13, 404.9, 404.91, 404.92, 404.93, 582, 582.1, 582.2, 582.4, 582.81, 582.89, 582.9, 583, 583.1, 583.2, 583.4, 583.6, 583.7, 583.81, 583.89, 583.9, 584.5, 584.6, 584.7, 584.8, 584.9, 585.1, 585.2, 585.3, 586, 587, 588, 588.81, 588.89, 588.9, 753, 753.12, 753.13, 753.14, 753.15, 753.16, 753.17, 753.19, 788.5, 792.5, V42.0, V45.11, V45.12, V56.0, V56.1, V56.2, V56.31, V56.32, V56.8, E08.22, E09.22, E10.22, E11.22, E13.22, I12.0, I12.9, I13.0, I13.1, I13.10, I13.11, I13.2, N18.1, N18.2, N18.3, N18.4, N18.5, N18.9, N19, N99.0, R34, Z49.01, Z49.02, Z49.31, Z49.32, Z99.2
Chronic Kidney Disease - Severe	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 585.4, 585.5, 585.6, 788.5, 792.5, V42.0, V45.11, V45.12, V56.0, V56.1, V56.2, V56.31, V56.32, V56.8, I12.0, I13.11, I13.2, N18.4, N18.5, N18.6, R34, Z99.2
Hyperlipidemia	272, 272.1, 272.2, 272.3, 272.4, 272.5, 272.6, 272.7, 272.8, 272.9, 759.9, E71.30, E75.21, E75.22, E75.5, E75.6, E77.0, E77.1, E78.0, E78.1, E78.2, E78.3, E78.4, E78.5, E78.6, E78.7, E78.70, E78.79, E78.81, E78.89, E78.9, E88.1, E88.89
Valvular Disease	35.05, 35.12, 35.10, 35.11, 35.14, 35.20, 35.21, 35.06, 35.13, 35.22, 35.24, 35.25, 35.26, 35.27, 35.28, 35.96, 35.23, 394.1, 394.2, 396.3, 396.2, 396.9, 394.9, 396.0, 396.1, 396.8, 394.0, V42.3, V43.3, I05.0, I05.1, I05.2, I05.8, I05.9, I06.8 , I06.9, I07.8, I07.9, I08.0, I08.1, I08.3, I08.8, I08.9, I09.1, I34.0, I34.1, I34.2, I34.8, I34.9, I35.0, I35.1, I35.2, I35.8, I35.9, I36.0, I36.1, I36.2, I36.8, I36.9, I37.0, I37.1, I37.2, I37.8, I37.9, I38
Prior Stroke / Transient Ischemic Attack	362.31, 362.32, 362.33, 362.34, 388.02, 430, 431, 432.9, 433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.00, 434.01, 434.10, 434.11, 434.91, 435.0, 435.1, 435.2,

	435.3, 435.8, 435.9, 437.1, 437.7, 437.9, 438.10, 438.11, 438.12, 438.13, 438.14, 438.20, 438.21, 438.22, 438.81, 438.82, 438.83, 438.89, 438.9, 997.02, V12.54, G45.0, G45.1, G45.2, G45.3, G45.4, G45.8, G46.3, G46.4, H34.00, H34.01, H34.02, H34.03, H34.10, H34.11, H34.12, H34.13, H34.211, H34.212, H34.213, H34.219, H34.231, H34.232, H34.233, H34.239, H93.099, I60.9, I61.9, I62.9, I63.00, I63.011, I63.012, I63.019, I63.111, I63.112, I63.119, I63.12, I63.131, I63.132, I63.139, I63.19, I63.20, I63.211, I63.212, I63.219, I63.22, I63.231, I63.232, I63.239, I63.29, I63.30, I63.311, I63.312, I63.319, I63.321, I63.322, I63.329, I63.331, I63.332, I63.339, I63.341, I63.342, I63.349, I63.40, I63.411, I63.412, I63.419, I63.421, I63.422, I63.429, I63.431, I63.432, I63.439, I63.49, I63.50, I63.511, I63.512, I63.519, I63.521, I63.522, I63.529, I63.531, I63.532, I63.539, I63.541, I63.542, I63.549, I63.59, I63.6, I63.8, I63.9, I66.01, I66.02, I66.03, I66.09, I66.11, I66.12, I66.13, I66.19, I66.21, I66.22, I66.23, I66.29, I66.3, I66.8, I66.9, I67.81, I67.82, I67.841, I67.848, I67.89, I67.9, I69.80, I69.81, I69.820, I69.821, I69.822, I69.823, I69.828, I69.831, I69.832, I69.833, I69.834, I69.839, I69.841, I69.842, I69.843, I69.844, I69.849, I69.851, I69.852, I69.853, I69.854, I69.859, I69.861, I69.862, I69.863, I69.864, I69.865, I69.869, I69.890, I69.891, I69.892, I69.893, I69.898, I69.90, I69.91, I69.920, I69.921, I69.922, I69.923, I69.928, I69.931, I69.932, I69.933, I69.934, I69.939, I69.941, I69.942, I69.943, I69.944, I69.949, I69.951, I69.952, I69.953, I69.954, I69.959, I69.961, I69.962, I69.963, I69.964, I69.965, I69.969, I69.990, I69.991, I69.992, I69.993, I69.998, I97.810, I97.811, I97.820, I97.821, Z86.73, G45.9
Systemic Atherosclerosis	441, 441.01, 441.02, 441.03, 441.1, 441.2, 441.3, 441.4, 441.5, 441.6, 441.7, 441.9, I70.0, I71.00, I71.01, I71.02, I71.03, I71.1, I71.2, I71.3, I71.4, I71.5, I71.6, I71.8, I71.9
Cerebral Atherosclerosis	433, 433.01, 433.1, 433.11, 433.2, 433.21, 433.3, 433.31, 433.8, 433.81, 433.91, 434.9, 434.91, 435, 435.1, 435.2, 435.3, 437, 437.1, 438.13, 438.14, G45.0, G45.8, I63.00, I63.011, I63.012, I63.019, I63.111, I63.112, I63.119, I63.12, I63.131, I63.132, I63.139, I63.19, I63.20, I63.211, I63.212, I63.219, I63.22, I63.231, I63.232, I63.239, I63.29, I63.30, I63.311, I63.312, I63.319, I63.321, I63.322, I63.329, I63.331, I63.332, I63.339, I63.341, I63.342, I63.349, I63.40, I63.411, I63.412,

	I63.419, I63.421, I63.422, I63.429, I63.431, I63.432, I63.439, I63.49, I63.50, I63.511, I63.512, I63.519, I63.521, I63.522, I63.529, I63.531, I63.532, I63.539, I63.541, I63.542, I63.549, I63.59, I65.01, I65.02, I65.03, I65.09, I65.1, I65.21, I65.22, I65.23, I65.29, I65.8, I65.9, I66.9, I67.2, I67.81, I67.82, I67.89
Thyrotoxicosis	242.00, 242.01, 242.10, 242.11, 242.20, 242.21, 242.30, 242.31, 242.40, 242.41, 242.80, 242.81, 242.90, 242.91, E05.00, E05.01, E05.10, E05.20, E05.21, E05.80, E05.41, E05.90, E05.30, E05.31, E05.40, E05.81, E05.11, E05.91
Hypothyroidism	243, 244, 244.1, 244.8, 244.9, 245, 245.1, 245.2, 245.9, E03.1, E03.8, E03.9, E06.1, E06.3, E06.5, E06.9, E89.0
Pulmonary Disease	490, 491.0, 491.1, 491.20, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 493.00, 493.01, 493.02, 493.10, 493.11, 493.12, 493.20, 493.21, 493.22, 493.81, 493.82, 493.90, 493.91, 493.92, 494.0, 494.1, 495.0, 495.1, 495.2, 495.3, 495.4, 495.5, 495.6, 495.7, 495.8, 495.9, 496, 500, 501, 502, 503, 504, 505, 506.0, 506.1, 506.2, 506.3, 506.4, 506.9, A15.0, A52.72, B38.1, B39.1, B40.1, D86.0, D86.2, E84.0, J40, J41.0, J41.1, J41.8, J42, J43.0, J43.1, J43.2, J43.8, J43.9, J44, J44.0, J44.1, J44.9, J45.20, J45.21, J45.22, J45.30, J45.31, J45.32, J45.40, J45.41, J45.42, J45.50, J45.51, J45.52, J45.901, J45.902, J45.909, J45.990, J45.991, J45.998, J47, J47.0, J47.1, J47.9, J60, J61, J62.0, J62.8, J63.0, J63.1, J63.2, J63.3, J63.4, J63.5, J63.6, J64, J65, J66.0, J66.1, J66.2, J66.8, J67.0, J67.2, J67.4, J67.5, J67.6, J67.7, J67.8, J67.9, J671, J673, J68.0, J68.1, J68.2, J68.3, J68.4, J68.9, J70.1, J70.3, J70.4, J81.8, J82, J84.02, J84.03, J84.10, J84.112, J84.115, J84.17, J84.82, J84.842, J84.89, J84.9, J95.3, J98.2, J98.3, M30.1, M32.13, M34.81, M35.02
Chronic Obstructive Pulmonary Disease	491, 491.0, 491.1, 491.2, 491.20, 491.21, 491.22, 491.8, 491.9, 492, 492.0, 492.8, 493.2, 493.20, 493.21, 493.22, 496, J44, J44.0, J44.1, J44.9, J41, J41.0, J41.1, J41.8, J42, J43, J43.1, J43.2, J43.8, J43.9, J45.5, J45.50, J45.51, J45.52
Congenital Heart Disease	745.0, 745.1, 745.2, 745.4, 745.5, 746.1, 746.2, 746.3, 746.4, 746.5, 746.6, 746.7, 746.81, 746.82, 746.83, 746.85, 746.86, 747.1, 747.11, 747.29, 747.31, 747.49, 745.7, 745.11, 745.3, 745.12, 745.8, 746.9, 745.69, 745.9, 746.01, 746.02, 746.09, 746.00, 746.1, 746.89, 747.6, 747.0, 747.9, 748.5, Q20.0, Q20.3, Q21.3, Q21.0, Q21.1, Q22.4, Q22.5, Q23.0, Q23.1, Q23.2, Q23.3, Q23.4, Q24.4, Q24.2, Q24.3, Q24.5, Q24.6, Q25.1, Q25.2, Q25.3, Q25.4, Q25.8, Q25.9, Q25.5, Q25.6,

	Q25.7, Q26, Q20.8, Q20.1, Q20.2, Q20.4, Q20.5, Q20.6, Q20.8, Q20.9, Q21.2, Q21.4, Q21.8, Q21.9, Q22.0, Q22.1, Q22.2, Q22.3, Q22.6, Q22.8, Q22.9, Q23.8, Q23.9, Q27.4, Q24.8, Q24.9, Q25.0, Q28.9, Q33.2
Cardiomegaly	429.3, I51.7
Mitral Valve Disorder	394, 394.0, 394.1, 394.2, 394.9, 424.0, I34, I34.0, I34.1, I34.2, I34.8, I34.9, I05, I05.0, I05.1, I05.2, I05.8, I05.9
Mitral Stenosis	394.0, 394.2, I05.0, I05.2, I34.2
Mitral Insufficiency	394.1, 394.2, I05.1, I34.0, I05.2
Mitral Valve Prolapse	I34.1, 424.0
Supraventricular Tachycardia	427.0, I47.1
Premature Atrial Contractions	427.61, I49.1
Alcohol Use – Heavy	303, 303.0, 303.00, 303.01, 303.02, 303.9, 303.90, 303.91, 303.92, 305, 305.00, 305.01, 305.02, F10.1, F10.10, F10.12, F10.120, F10.121, F10.129, F10.14, F10.15, F10.150, F10.151, F10.159, F10.18, F10.180, F10.181, F10.182, F10.188, F10.19, F10.2, F10.20, F10.22, F10.220, F10.221, F10.229, F10.23, F10.230, F10.231, F10.232, F10.239, F10.24, F10.25, F10.250, F10.251, F10.259, F10.26, F10.27, F10.28, F10.280, F10.281, F10.282, F10.288, F10.9, F10.92, F10.920, F10.921, F10.929, F10.94, F10.95, F10.950, F10.951, F10.959, F10.96, F10.97, F10.98, F10.980, F10.981, F10.982, F10.988, F10.99
Pericarditis	036.41, 074.21, 093.81, 420, 098.83, 420.9, 420.99, 423.1, 423.2, 393, 420.91, 420.9, 115.93, 391.0, 115.03, 115.13, 115.93, 423.0, 423.3, 423.8, 423.9, 420.0, A39.53, B33.23, I01.0, I09.2, I30.0, I30.1, I30.8, I30.9, I31.0, I31.1, I31.2, I31.3, I31.4, I31.8, I31.9, I32
Myocarditis	42.9, 032.82, 036.43, 074.23, 093.82, 13.03, 391.2, 398.0, 422.0, 422.90, 422.91, 422.92, 422.93, 422.99, I40.0, I40.1, I40.8, I40.9, I41, I51.4, J10.82, J11.82, A38.1, A39.52, B26.82, B33.22, B58.81, D86.85, I01.2, I09.0
Obstructive Sleep Apnea	780.57, 327.23, G47.30, G47.33, G47.39
Prior Cardiac Surgery	35.00, 35.01, 35.02, 35.03, 35.04, 35.11, 35.22, 35.31, 35.32, 35.33, 35.34, 35.35, 35.39, 35.41, 35.42, 35.5, 35.51, 35.52, 35.53, 35.54, 35.55, 35.6, 35.61, 35.62, 35.63, 35.7, 35.71, 35.72, 35.73, 35.8, 36.03, 36.04, 36.07, 36.09, 36.10, 36.1, 36.11, 36.12, 36.13, 36.14, 36.15, 36.16, 36.17, 36.19, 36.2, 37.0, 37.1, 37.11, 37.12, 37.2, 37.21, 37.22, 37.23, 37.24, 37.25, 37.26, 37.27, 37.28, 37.29, 37.3, 37.31, 37.32, 37.33, 37.34, 37.35, 37.36, 37.37, 37.4, 37.41, 37.49, 37.7, 37.8, 37.9, I97.190, Z98.61, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721,

	I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I97.0, I97.110, I97.120, I97.130, V45.82, Z45.018, Z45.02, Z45.09, Z48.21, Z48.280, Z95.1, Z95.2, Z95.811, Z95.812
Hypertrophic Cardiomyopathy	425.1, 425.11, 425.18, I42.1, I42.2
Other Cardiomyopathy	425, 425.0, 425.2, 425.3, 425.4, 425.5, 425.7, 425.8, 425.9, I42, I42.0, I42.3, I42.4, I42.5, I42.6, I42.7, I42.8, I42.9
Chronic Liver Disease	070.0, 070.20, 070.21, 070.22, 070.23, 070.30, 070.31, 070.32, 070.33, 070.41, 070.42, 070.43, 070.49, 070.51, 070.52, 070.53, 070.54, 070.59, 070.6, 070.70, 070.71, 070.9, 273.4, 275.01, 275.1, 453.0, 570, 571.1, 571.3, 571.40, 571.41, 571.49, 571.8, 571.9, 573.0, 573.1, 573.2, 573.3, 573.4, 573.8, 573.9, 576.1, B15.0, B15.9, B16.0, B16.1, B16.2, B16.9, B17.0, B17.10, B17.11, B17.2, B17.8, B17.9, B18.0, B18.1, B18.2, B18.8, B18.9, B19.0, B19.10, B19.11, B19.20, B19.21, B19.9, B94.2, K70.0, K70.10, K70.11, K70.2, K70.30, K70.31, K70.9, K71.0, K71.10, K71.11, K71.2, K71.3, K71.4, K71.50, K71.51, K71.6, K71.7, K73.0, K73.1, K73.2, K73.8, K73.9, K74.0, K74.1, K75.3, K75.4, K75.8, K75.89, K75.9, K76.89, Z22.50, Z22.51, Z22.52, Z22.59, K72.00, K72.01, K72.1, K72.10, K72.22, K72.90, K72.91
Cirrhosis	571.2, 571.5, 571.6, K70.40, K70.41, K71.9, K74.2, K74.3, K74.5, K74.60, K74.69, K744, K7460, K75.0, K75.2
Liver Complications	456.0, 456.20, 456.1, 456.21, 572.2, 572.3, 572.4, 572.8, 573.5, 567.0, 567.23, 567.21, 567.29, 567.1, 567.89, 567.9, 789.5, 789.59, K76.0, K76.1, K76.2, K76.3, I85.01, I85.11, I85, K76.6, K76.7, K76.81, K76.80, K76.8, K65.0, K65.2, K65.8, K65.9, R18, R18.0, R18.8

ICD: International Classification of Disease

Table S3. CHARGE-AF score components and weights*

Covariate	Estimated β (SE)
Age (per 5-year increase)	0.508 (0.022)
Race (white)	0.465 (0.093)
Height (per 10 cm increase), cm	0.248 (0.036)
Weight (per 15 kg increase), kg	0.115 (0.033)
Systolic blood pressure (per 20 mmHg increase), mmHg	0.197 (0.033)
Diastolic blood pressure (per 10 mmHg increase), mmHg	-0.101 (0.032)
Current smoker	0.359 (0.091)
Anti-hypertensive medication use	0.349 (0.063)
Diabetes	0.237 (0.073)
Heart failure	0.701 (0.106)
Myocardial infarction	0.496 (0.089)

CHARGE-AF: Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation; SE: standard error

* Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2(2):e000102.

Table S4. Concept unique identifiers (CUIs) mapped to potential predictors from linkage to the Unified Medical Language System (UMLS)*

Variable	CUI
Alcohol Abuse	C0085762
Alcohol Abuse	C0001973
Alcohol Abuse	C0560219
Cardiomegaly	C0018800
Cerebral Atherosclerosis	C4024924
Cerebrovascular disease	C0007775
Cerebrovascular disease	C0007820
Chronic Kidney Disease - Severe	C2316810
Chronic Kidney Disease	C1561643
Chronic Liver Disease	C0341439
Chronic Liver Disease	C0085605
Chronic Obstructive Pulmonary Disease	C0024117
Chronic Obstructive Pulmonary Disease	C0034067
Chronic Obstructive Pulmonary Disease	C0008677
Cirrhosis	C0023890
Congenital Heart Disease	C0152021
Congestive heart failure	C0018802
Coronary artery disease	C0010054
Coronary artery disease	C0010068
Diabetes	C0011849
Diabetes	C0011860
Diabetes	C0011854
Hyperlipidemia	C0020473
Hypertension	C0020538
Hypertrophic Cardiomyopathy	C0007194
Hypertrophic Cardiomyopathy	C4551472
Hypothyroidism	C0020676
Left atrial enlargement	C0232309
Left atrial enlargement	C0232310
Left atrial enlargement	C0238705
Left Ventricular Hypertrophy	C0149721
Left Ventricular Hypertrophy	C0232306
Left Ventricular Hypertrophy	C0344398
Liver Complications	C0015695
Liver Complications	C0267821
Liver Complications	C0014867
Liver Complications	C0019151
Liver Complications	C0020541
Liver Complications	C0019212
Liver Complications	C0600452

Mitral Insufficiency	C0026266
Mitral Stenosis	C0026269
Mitral Valve Disease	C0026265
Mitral Valve Prolapse	C0026267
Myocardial Infarction	C0027051
Myocarditis	C0027059
Obesity	C0028754
Obesity	C0028756
Obstructive Sleep Apnea	C0520679
Other Cardiomyopathy	C0878544
Other Cardiomyopathy	C0007193
Other Cardiomyopathy	C0007192
Other Cardiomyopathy	C0264834
Pericarditis	C0031046
Peripheral vascular disease	C0085096
Premature atrial contractions	C0033036
Prolonged PR Interval	C0600125
Pulmonary Disease	C0024115
Shortened PR Interval	C0520878
Stroke	C0038454
Supraventricular Tachycardia	C0039240
Supraventricular Tachycardia	C1963244
Supraventricular Tachycardia	C3815188
Supraventricular Tachycardia	C0030590
Systemic Atherosclerosis	C0155733
Thyrotoxicosis	C0040156
Thyrotoxicosis	C0020550
Transient Ischemic Attack	C0007787
Valvular Disease	C3258293
Valvular Disease	C0018824

CUI: concept unique identifier; UMLS: Unified Medical Language System

* Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004;32:D267-270.

Table S5. C-statistics and hazard ratios in prediction models that exclude race and include social determinants of health

	C-statistic (95% CI)	Hazard Ratio (95% CI)
Development (Add insurance)		
Codified+NLP	0.742 (0.733-0.751)	
≥ 84th percentile		15.91 (12.88-19.64)
50-<84th percentile		5.74 (4.65-7.10)
16-<50th percentile		2.32 (1.86-2.90)
< 16th percentile		-
Codified-only	0.728 (0.719-0.737)	
≥ 84th percentile		12.80 (10.55-15.52)
50-<84th percentile		4.64 (3.82-5.63)
16-<50th percentile		2.17 (1.77-2.66)
< 16th percentile		-
Development (Add insurance + Income)*		
Codified+NLP	0.741 (0.732-0.750)	
≥ 84th percentile		15.44 (12.52-19.04)
50-<84th percentile		5.81 (4.71-7.17)
16-<50th percentile		2.29 (1.83-2.86)
< 16th percentile		-
Codified-only	0.727 (0.717-0.736)	
≥ 84th percentile		12.59 (10.33-15.36)
50-<84th percentile		4.68 (3.84-5.72)
16-<50th percentile		2.13 (1.73-2.63)
< 16th percentile		-
Internal Validation (Add insurance)		
Codified+NLP	0.731 (0.716-0.746)	
≥ 84th percentile		12.91 (9.42-17.71)
50-<84th percentile		4.38 (3.18-6.02)
16-<50th percentile		2.00 (1.43-2.80)
< 16th percentile		-
Codified-only	0.722 (0.707-0.737)	
≥ 84th percentile		12.02 (8.76-16.50)
50-<84th percentile		4.74 (3.45-6.52)
16-<50th percentile		1.96 (1.40-2.74)
< 16th percentile		-
Internal Validation (Add insurance + Income)*		
Codified+NLP	0.731 (0.716-0.746)	
≥ 84th percentile		12.72 (9.27-17.46)
50-<84th percentile		4.26 (3.09-5.86)

16-<50th percentile		1.95 (1.39-2.73)
< 16th percentile		-
Codified-only	0.722 (0.707-0.738)	
≥ 84th percentile		11.36 (8.27-15.60)
50-<84th percentile		4.55 (3.31-6.25)
16-<50th percentile		1.99 (1.42-2.78)
< 16th percentile		-

NLP: natural language processing; CI: confidence interval

* Income represents proportion of population by zip code with income < $50,000 from 2008-2012 ascertained from Melendez, Robert, Clarke, Philippa, Khan, Anam, Gomez-Lopez, Iris, Li, Mao, and Chenoweth, Megan. National Neighborhood Data Archive (NaNDA): Socioeconomic Status and Demographic Characteristics of ZIP Code Tabulation Areas, United States, 2008-2017. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-07-30. https://doi.org/10.3886/E120462V1. Missing income data in development population: n=1,498; Missing income data in internal validation population: n=507

Table S6. Hazard ratios and 95% confidence intervals for incidence of AF by risk groups defined by the 16th, 50th, and 84th percentiles for each model in the external validation cohort

	Codified+NLP	Codified-only	CHARGE-AF
	HR (95% CI)	HR (95% CI)	HR (95% CI)
≥ 84th percentile	17.74 (13.59-23.16)	13.26 (10.47-16.78)	14.60 (11.37-18.77)
50-<84th percentile	5.67 (4.34-7.43)	4.26 (3.36-5.41)	5.24 (4.07-6.74)
16-<50th percentile	2.41 (1.82-3.20)	2.01 (1.56-2.58)	2.23 (1.71-2.90)
< 16th percentile	-	-	-

AF: atrial fibrillation; NLP: natural language processing; CI: confidence interval

Table S7. Percentile-based net reclassification improvement (NRI) with groups determined by 16th, 50th, 84th percentile of each model in external validation cohort

	Overall NRI (95% CI)	Event NRI (95% CI)	Non-event NRI (95% CI)
Codified+NLP vs. CHARGE-AF	0.044 (0.018 - 0.069)	0.040 (0.016 – 0.064)	0.003 (-0.003 – 0.010)
Codified-only vs. CHARGE-AF	0.001 (-0.025 - 0.026)	-0.003 (-0.026 – 0.021)	0.004 (-0.002 – 0.011)
Codified+NLP vs. Codified-only	0.051 (0.027 - 0.072)	0.055 (0.033 – 0.075)	-0.004 (-0.010 – 0.002)

NRI: net reclassification improvement; NLP: natural language processing; CI: confidence interval; CHARGE-AF: Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation

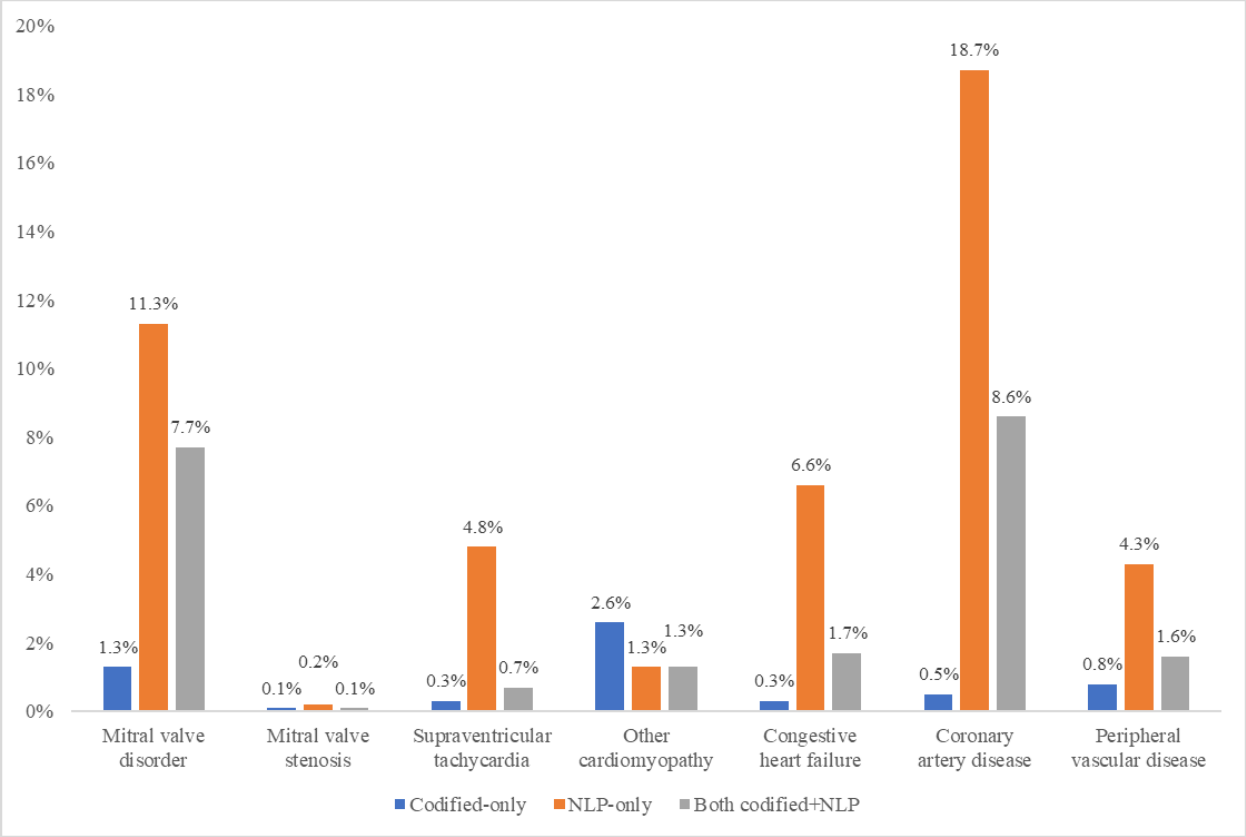
Table S8. C-statistics and hazard ratios in external validation cohort over 3-years of follow-up

	C-statistic (95% CI)	Hazard Ratio (95% CI)
Codified+NLP	0.758 (0.746-0.769)*	
≥ 84th percentile		19.36 (14.29-26.21)
50-<84th percentile		5.93 (4.37)
16-<50th percentile		2.39 (1.73-3.30)
< 16th percentile		-
Codified-only	0.745 (0.733-0.757)	
≥ 84th percentile		13.22 (10.15-17.23)
50-<84th percentile		4.35 (3.33-5.69)
16-<50th percentile		1.79 (1.34-2.38)
< 16th percentile		-
CHARGE-AF	0.741 (0.730-0.753)	
≥ 84th percentile		16.89 (12.51-22.82)
50-<84th percentile		6.00 (4.43-8.12)
16-<50th percentile		2.44 (1.78-3.36)
< 16th percentile		-

CI: confidence interval, NLP: natural language processing; CHARGE-AF: Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation

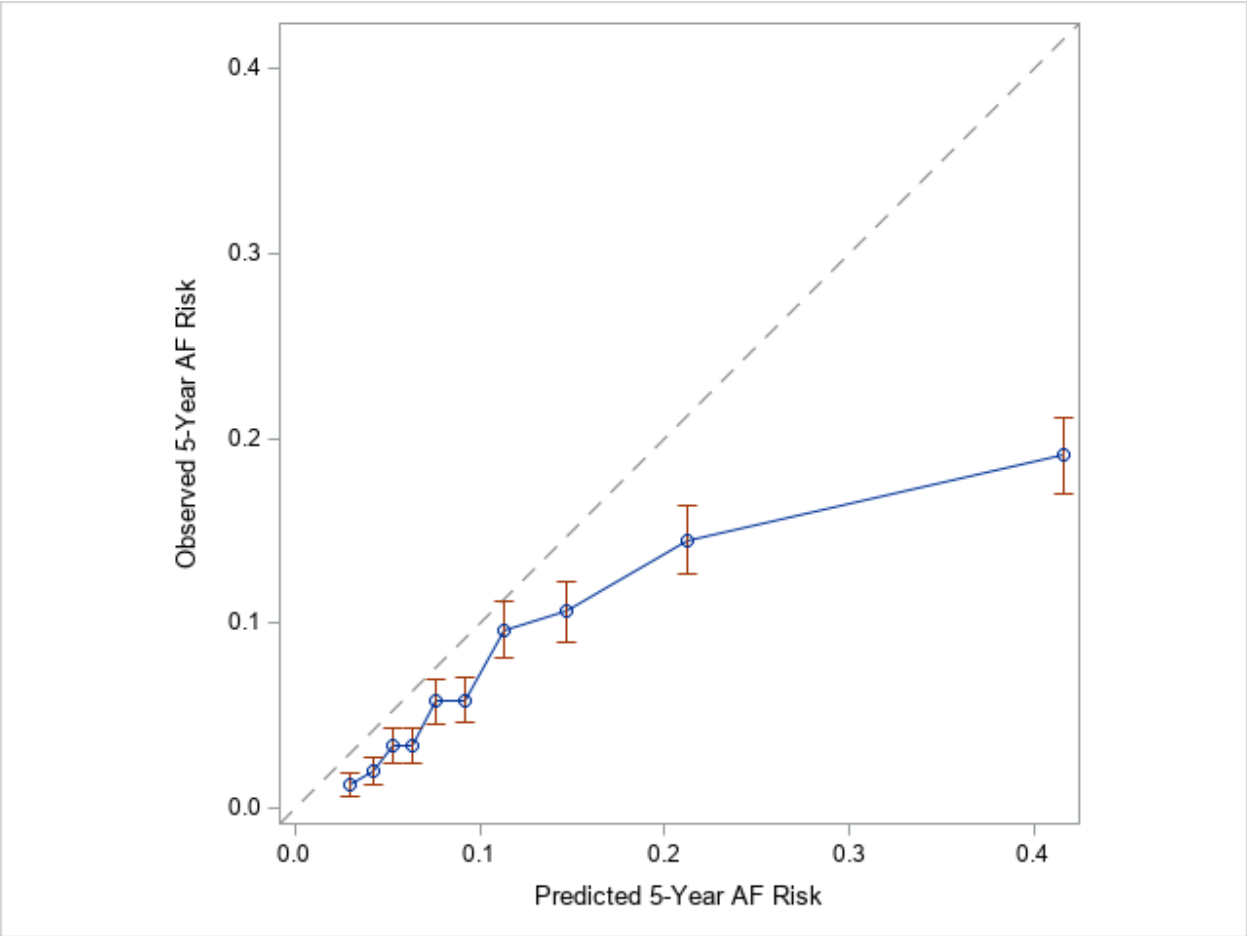
* $p<0.001$ comparing C-statistic in Codified+NLP compared to codified-only and CHARGE-AF

Figure S1. Prevalence of features identified by codified data only, NLP data only, and by both codified and NLP data among those where both the codified and NLP version were selected for inclusion in the codified+NLP model in the development cohort



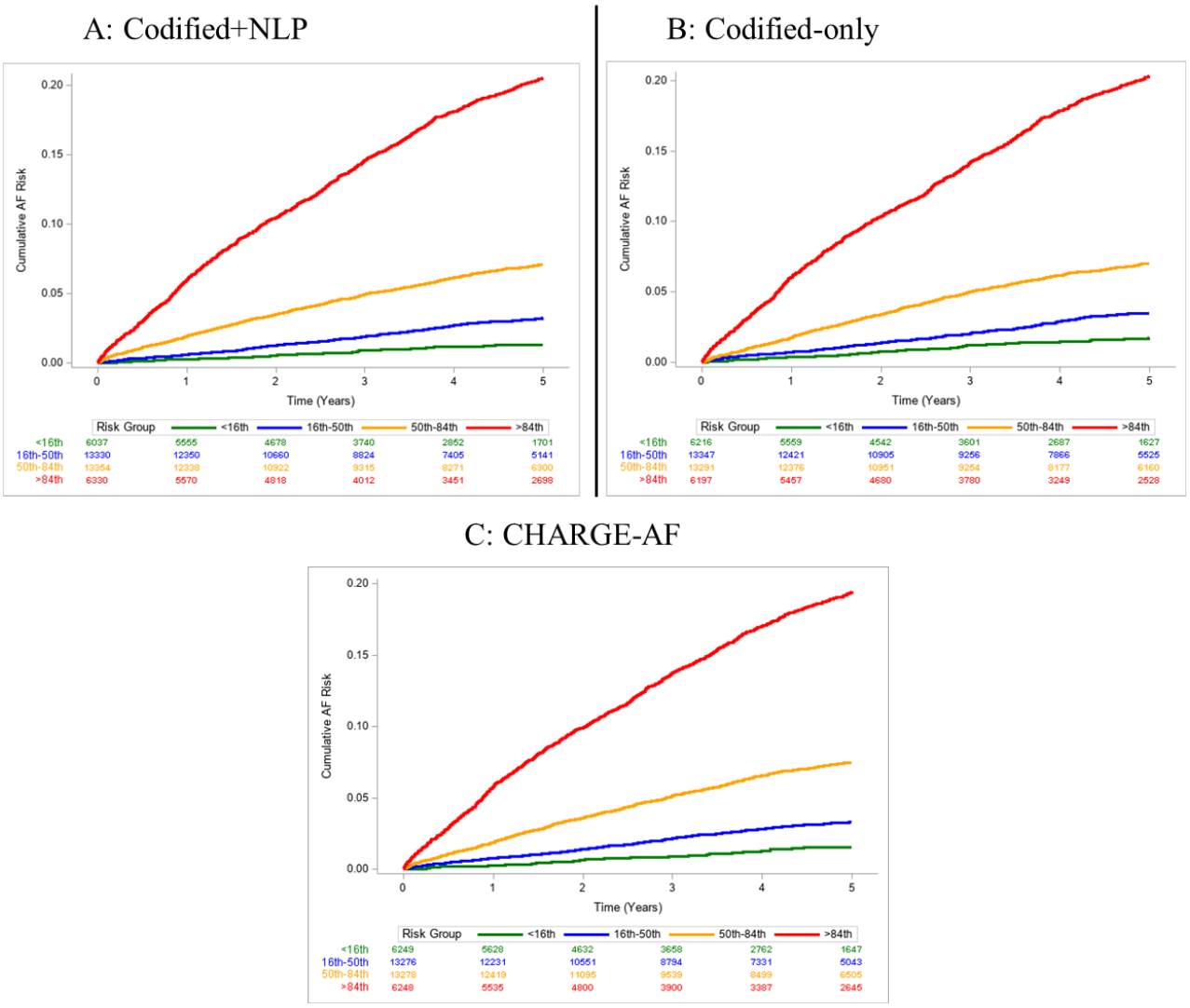
NLP: natural language processing

Figure S2. Plot of observed 5-year AF risk versus predicted 5-year AF risk for recalibrated CHARGE-AF with patients divided into risk groups based on deciles in the internal validation cohort



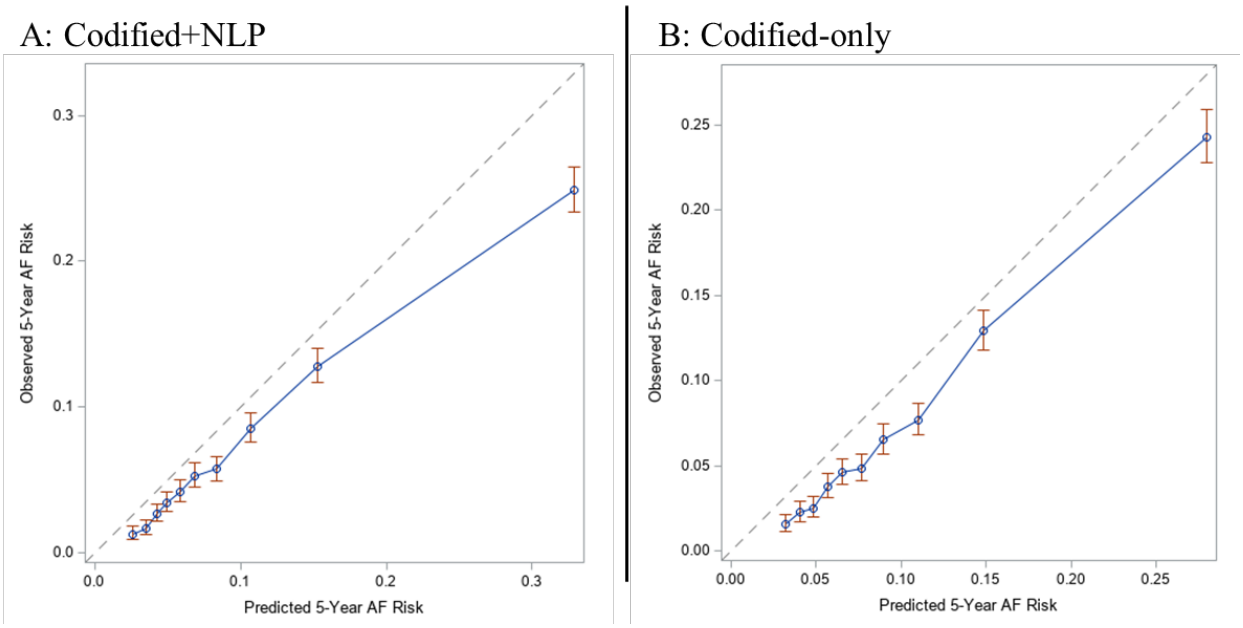
AF: atrial fibrillation; CHARGE-AF: Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation

Figure S3. Cumulative incidence plots stratified by groups of predicted risk for codified + NLP, codified-only, and CHARGE-AF models in external validation cohort



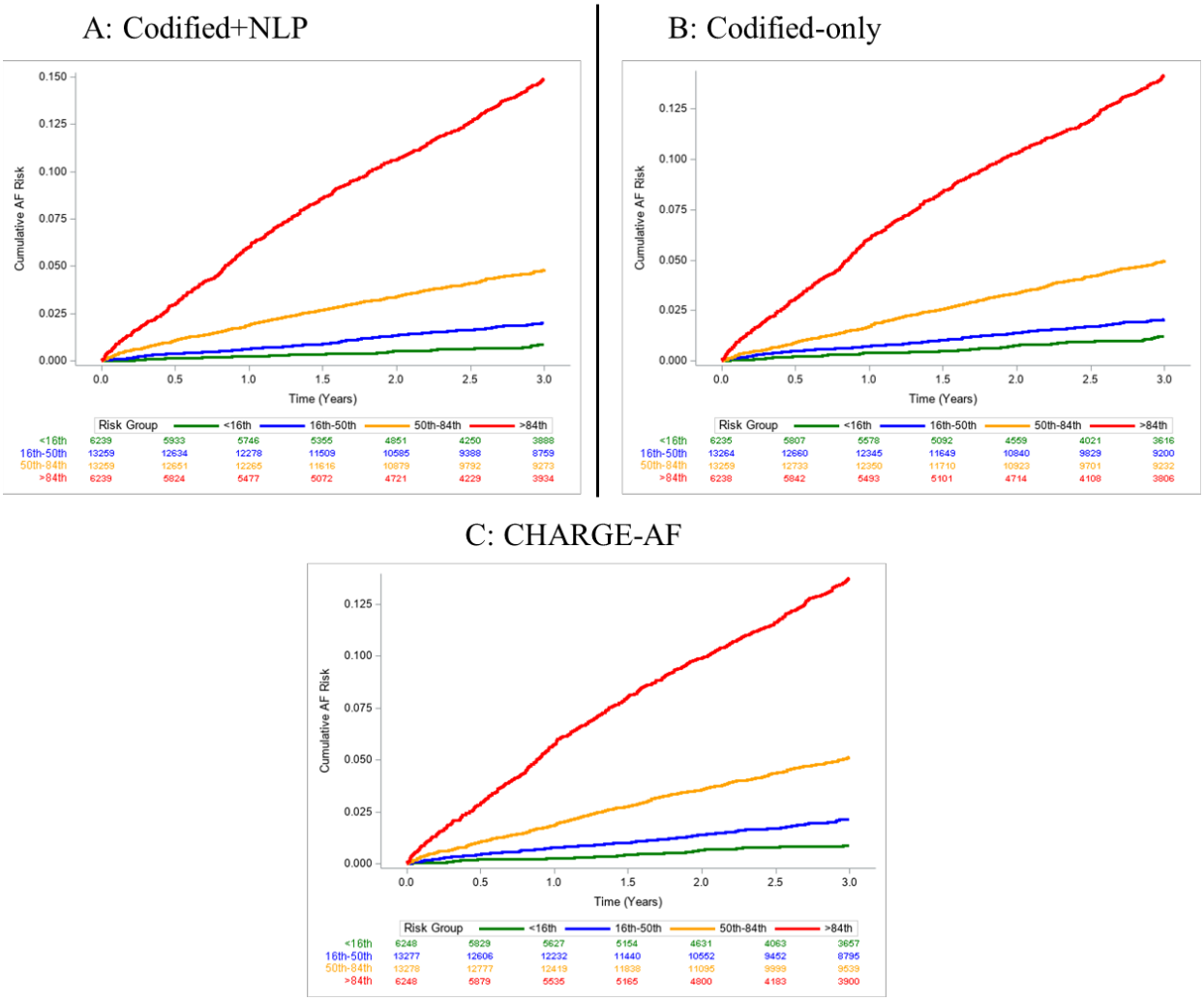
NLP: natural language processing; AF: atrial fibrillation; CHARGE-AF: Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation

Figure S4. Plots of observed 5-year AF risk versus predicted 5-year AF risk with patients divided into risk groups based on deciles in external validation cohort



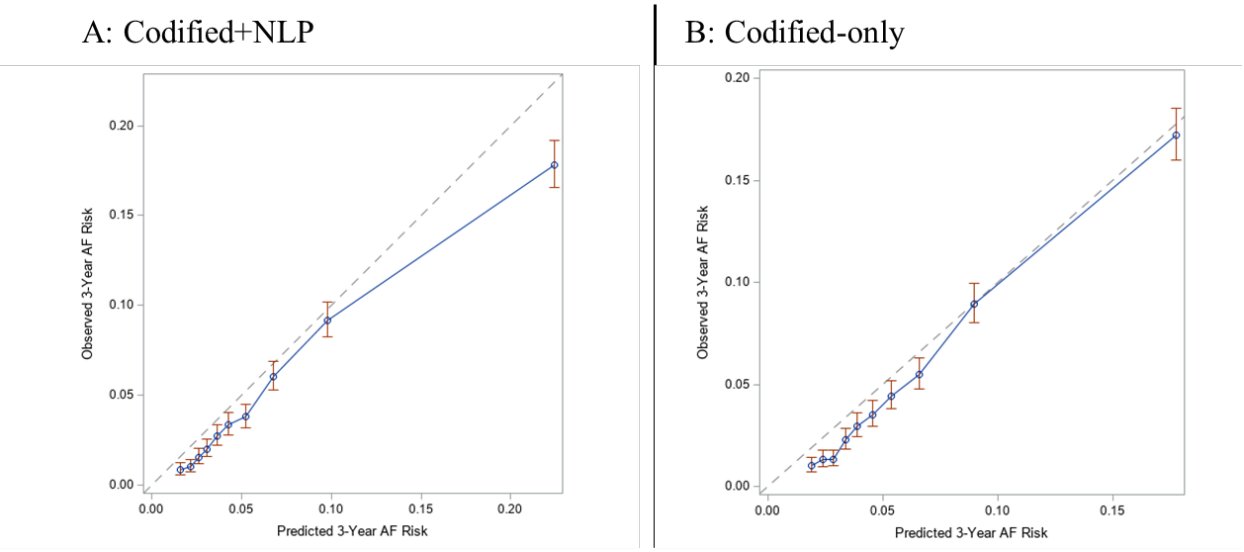
AF: atrial fibrillation; NLP: natural language processing

Figure S5. Cumulative incidence plots stratified by groups of predicted risk for codified + NLP, codified-only, and CHARGE-AF models in external validation cohort over 3-years of follow-up



NLP: natural language processing; CHARGE-AF: Cohorts for Heart and Aging Research in Genomic Epidemiology Atrial Fibrillation; AF: atrial fibrillation

Figure S6. Plots of observed 3-year AF risk versus predicted 3-year AF risk with patients divided into risk groups based on deciles in external validation cohort



AF: atrial fibrillation; NLP: natural language processing