Name:	Joel N. Swerdel, MS MPH
Affiliation:	Janssen Research and Development
Email:	jswerdel@its.jnj.com
Presentation type	Poster

Predictive Modeling of Incident Heart Failure in Subjects with Newly Diagnosed Atrial Fibrillation.

Joel N. Swerdel, MS MPH[1,3]; Jenna M. Reps, PhD[1,3]; Patrick B. Ryan, PhD[1,2,3]

[1] Janssen Research and Development, Titusville, NJ, USA

[2] Columbia University, New York, NY, USA

[3] Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Abstract

Background: The incidence of both atrial fibrillation (AF) and heart failure (HF) have been increasing over the past 20 years in the US. Negative outcomes including stroke, myocardial infarction, and death are experienced at a higher rate in those with both conditions compared to those with only one of the conditions. It is therefore important to understand the risk factors that make patients with one of the disorders prone to the development of the follow-up disease. The objective of this study was to use machine learning to develop a model for predicting the probability of developing HF in patients with newly diagnosed AF. Method: We used administrative claims data from the OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN) and Truven Medicare, and Commercial Claims and Encounters (CCAE) datasets from 2000 to 2016. The cohorts were developed using the OHDSI Atlas tool. The target populations were those patients with newly diagnosed AF and the outcome populations were those patients who developed HF at 3-6 months, 6-18 months, or 18-36 months after AF. In our models, we included covariates for condition occurrence, drug exposure, and clinical observations and measurements within 365 days of the index date (AF). Results: The areas under the Receiver Operating Curves were 0.73, 0.73, and 0.71 for developing HF at 3-6 months, 6-18 months, and 18-36 months after AF, respectively, in the Optum dataset indicating fair to good discrimination. The model predicted associations with HF development for many known risk factors such as cardiomyopathy, diabetes, and heart valve disorder. Conclusion: The models developed using the OHDSI PatientLevelPrediction package appear to show promise in predicting patients with AF at risk for developing HF between 3 months and 3 years after initial AF diagnosis.

Introduction

The incidence of both atrial fibrillation (AF) and heart failure (HF) have been increasing over the past 20 years in the US. It is estimated that by 2030 in the US, there will be 12 million people with AF and over 8 million with heart failure.[1, 2] Many studies have demonstrated a relationship between the 2 diseases; those with either one of the conditions are at a higher risk to develop the other condition than those without the condition. Negative outcomes including stroke, myocardial infarction, and death are experienced at a higher rate in those with both conditions compared to those with only one of the conditions. It is therefore important to understand the risk factors that make patients with one of the disorders prone to the development of the follow-up disease for reducing the risk of further negative outcomes. The objective of this study was to use machine learning to develop a model for predicting the probability of developing HF in patients with newly diagnosed AF.

Methods

Data for this study was collected between January 1, 2000 and December 31, 2016 from 4 data sets: Truven MarketScan Commercial Claims and Encounters (CCAE), Truven Medicare (MDCR), and OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN) (Optum). The databases had been translated from their original form into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The cohorts were developed using the OHDSI Atlas tool.

The cohort definitions used were as follows:

1) **Target population: Patients with newly diagnosed AF**
- A condition code for AF for the first time in a person's history (index date)
- A follow-up condition code for AF between 1 and 60 days following the index date
- No condition codes for HF in the person's history prior to the index date
- At least one in-patient or out-patient visit occurrence between 1 and 180 days prior to index date and at least one visit occurrence between 181 and 365 days prior to the index date
- A continuous observation period of at least 365 days prior to the index date

2) **Outcome population: Patients with newly diagnosed HF**
- A condition code for HF for the first time in a person's history
- A follow-up condition code for HF between 1 and 60 days following the initial HF diagnosis

We developed models using the Optum dataset at 3 different times-at-risk for developing HF: 3-6 months, 6-18 months, or 18-36 months after the AF index date. For each time-at-risk, we required a minimum continuous observation period after the index date up to the start of the period, e.g., for the 3-6 month time-at-risk window, patients eligibility required having at least 3 months continuous observation after the index date. Our datasets included all subjects who developed the outcome and a 10% random sample of those who did not develop the outcome. For these analyses, the R PatientLevelPrediction package was used. In our models, we included covariates for condition occurrence, drug exposure, and clinical observations and measurements within 365 days of the index date. LASSO (L1) logistic regression models were trained on 75% of the cohort records and tested on the remaining 25%. Internal validation of model performance was through analysis of the area under the Receiver Operator Characteristic (ROC) curves (AUC) and calibration of the Optum test data set. External validation of model performance was through analysis of the area under the Receiver Operator Characteristic (ROC) curves (AUC) of the CCAE and MDCR datasets.

Results

For internal validation on the Optum test data, the AUCs (95% Confidence Intervals), calibration intercepts, and calibration slopes were 0.73 (0.71, 0.75), 0.00, 1.04 for the model at 3-6 months time-at-risk; 0.73 (0.71, 0.74), 0.01, 0.99 at 6-18 months; and 0.71 (0.69, 0.73), 0.00, 0.99 at 18-36 months, indicating fair to good performance. For external validation, in CCAE, the AUCs were 0.68 for the model at 3-6 months time-at-risk; 0.67 at 6-18 months; and 0.66 at 18-36 months, indicating fair performance. In the MDCR dataset, the AUCs, calibration intercepts, and calibration slopes were 0.64 for the model at 3-6 months time-at-risk; 0.62 at 6-18 months; and 0.60 at 18-36 months, indicating fair performance. Table 1 shows comparisons in factor prevalence between those who developed HF after AF and those who did not. The results from the model indicated associations to HF development in AF patients with many previously recognized conditions including cardiomyopathy, diabetes, and heart valve disorder.

Table 1: Demographic information and model-predicted factors with associations to heart failure and their proportions in those who developed the outcome (AF → HF) and those who did not (No HF) during the 6-18 month time-at-risk period.

	Optum		CCAE		MDCR	
	181-548D		181-548D		181-548D	
Name	AF -> HF	No HF	AF -> HF	No HF	AF -> HF	No HF
Participants (N(%))	4550 (100.0)	22424 (100.0)	1320 (100.0)	19778 (100.0)	5410 (100.0)	22693 (100.0)
Age (Mean(SD))	75.3 (9.2)	67.4 (12.8)	57.9 (6.0)	55.5 (8.6)	80.0 (7.4)	77.5 (7.2)
Age (Median(IQR))	76.9 (70.4-82.0)	69.2 (59.4-77.1)	59.4 (55.2-62.2)	57.9 (51.6-61.8)	80.1 (74.3-85.4)	77.1 (71.7-82.7)
Female (N(%))	2258 (49.6)	9525 (42.5)	413 (31.3)	6714 (33.9)	2869 (53.0)	11204 (49.4)
Number of Visits Past 365D (Median(IQR))	11.0 (6.0-19.0)	10.0 (6.0-17.0)	9.5 (5.0-17.0)	8.0 (5.0-14.0)	10.0 (6.0-17.0)	9.0 (5.0-15.0)
Charlson Index (Mean(SD))	3.9 (2.6)	3.3 (2.4)	3.3 (2.5)	2.5 (2.0)	3.6 (2.4)	3.3 (2.3)
DCSI (Mean(SD))	3.7 (1.8)	3.1 (1.5)	3.0 (1.5)	2.6 (1.1)	3.5 (1.7)	3.2 (1.5)
CHADS2VASc (Mean(SD))	3.9 (1.5)	3.1 (1.6)	2.1 (1.0)	1.8 (0.9)	4.0 (1.4)	3.6 (1.4)
Anemia (N(%))	357 (7.8)	1053 (4.7)	36 (2.7)	315 (1.6)	309 (5.7)	809 (3.6)
Arthritis (N(%))	504 (11.1)	2205 (9.8)	118 (8.9)	1476 (7.5)	469 (8.7)	1861 (8.2)
Cardiomyopathy (N(%))	156 (3.4)	494 (2.2)	70 (5.3)	357 (1.8)	141 (2.6)	356 (1.6)
Chronic Kidney Disease (N(%))	577 (12.7)	1455 (6.5)	83 (6.3)	423 (2.1)	445 (8.2)	1193 (5.3)
Congenital Heart Disease (N(%))	73 (1.6)	464 (2.1)	23 (1.7)	251 (1.3)	51 (0.9)	196 (0.9)
COPD (N(%))	825 (18.1)	2387 (10.6)	133 (10.1)	903 (4.6)	825 (15.2)	2460 (10.8)
Coronary Artery Disease (N(%))	1260 (27.7)	4447 (19.8)	274 (20.8)	2373 (12.0)	1315 (24.3)	4513 (19.9)
Type 1 Diabetes (N(%))	160 (3.5)	405 (1.8)	39 (3.0)	271 (1.4)	133 (2.5)	357 (1.6)
Type 2 Diabetes (N(%))	1514 (33.3)	5312 (23.7)	452 (34.2)	3596 (18.2)	1395 (25.8)	4722 (20.8)
Dyspnea (N(%))	1041 (22.9)	4462 (19.9)	308 (23.3)	3173 (16.0)	1039 (19.2)	3792 (16.7)
Edema (N(%))	648 (14.2)	1929 (8.6)	116 (8.8)	950 (4.8)	481 (8.9)	1564 (6.9)
Heart valve disorder (N(%))	918 (20.2)	4157 (18.5)	226 (17.1)	2551 (12.9)	1016 (18.8)	3671 (16.2)
Hypertension (N(%))	3647 (80.2)	15343 (68.4)	801 (60.7)	9825 (49.7)	3274 (60.5)	13508 (59.5)
Ischemic Heart Disease (N(%))	778 (17.1)	3007 (13.4)	199 (15.1)	1727 (8.7)	726 (13.4)	2702 (11.9)
AMI (N(%))	406 (8.9)	1405 (6.3)	101 (7.7)	720 (3.6)	325 (6.0)	1159 (5.1)
Peripheral nerve disease (N(%))	601 (13.2)	2384 (10.6)	133 (10.1)	1703 (8.6)	487 (9.0)	1918 (8.5)
Peripheral vascular disease (N(%))	1383 (30.4)	4818 (21.5)	221 (16.7)	2352 (11.9)	1410 (26.1)	4800 (21.2)
Stroke/TIA (N(%))	454 (10.0)	1552 (6.9)	55 (4.2)	620 (3.1)	471 (8.7)	1733 (7.6)

Conclusion

The models developed using the OHDSI PatientLevelPrediction package appear to show promise in predicting patients with AF at risk for developing HF between 3 months and 3 years after initial AF diagnosis. Future work in this area would include further examination of possible latent risk factors for the development of HF in patients with AF.

1. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail. 2013 May;6(3):606-19.
2. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of atrial fibrillation in the U.S. adult population. Am J Cardiol. 2013 Oct 15;112(8):1142-7.