Phenotype Development and Evaluation of Heart Failure: A Case Study in using Patient-Level Prediction to Improve Phenotype Validity

Pooja M. Desai, Anna Ostropolets, Lauren R. Richter, Harry Reyes Nieva, Matthew Spotnitz, Victor A. Rodriguez, Tony Y. Sun, Karthik Natarajan

Background

Phenotyping algorithms are commonly utilized in observational health research to define a cohort of subjects who satisfy one or more criteria for a duration of time.[1,2] Establishing robust phenotypes with high fidelity is often challenging.[3] The OHDSI community has developed a number of publicly available resources including the ATLAS interface, Phenotype Library, and computational packages to facilitate phenotype construction, evaluation, and dissemination.

The proposed OHDSI phenotype development and evaluation pipeline comprises a literature review, using PHOEBE for identifying relevant concept sets, cohort construction in ATLAS, and phenotype review *via* cohort characterization.[4] While this approach may be time-consuming, it is often necessary for rigorous phenotype definitions. Robust phenotype development may be especially challenging when attempting to isolate incident cases (*i.e.*, new-onset) of a chronic condition. Index date misspecification, when the index date of record for an event differs from its actual onset date, is of particular concern when working with observational health data.

In constructing a heart failure (HF) phenotype, our initial aim was to predict the risk of HF-related hospitalization among thiazolidine (TZD) monotherapy-exposed patients with or without a history of HF. TZDs are a potent class of anti-diabetic drugs. Prior clinical studies have indicated that TZD use in vulnerable patients can increase the risk of HF exacerbations, even in those without a known diagnosis of HF.[5] The objective of this case study is to present an example of how tools ostensibly intended for tasks performed after phenotype evaluation, such as the PatientLevelPrediction (PLP) package, may be leveraged to test, validate, and improve phenotype definitions.

Methods

Phenotype development started with literature review concerning TZD monotherapy and clinical definitions of HF-related codes using standard terminologies (*e.g.*, SNOMED). Pre-existing phenotypes within the OHDSI Phenotype library were also reviewed on data.ohdsi.org. We constructed an ATLAS cohort defining HF as any one of the following: (1) a condition occurrence of HF diagnostic codes (SNOMED 84114007 Heart Failure and all descendants); (2) an observation of other HF-related condition codes (*e.g.*, systolic or diastolic dysfunction); (3) a condition occurrence of other HF-related codes (such as DRG codes of HF and shock), or; (4) an echocardiogram ejection fraction measurement with a value of less than 50%. Our outcome cohort was TZD monotherapy-exposed patients with HF-related hospitalization, defined as an inpatient or emergency room-inpatient visit during which an incident of HF (as defined above) occurred. Detailed specifications are available online as an OHDSI ATLAS Cohort Definition.[1]

After defining the HF cohort, we ran characterization in ATLAS to assess demographic, drug, and condition covariates. We used data from the Columbia University Irving Medical Center (CUIMC) electronic health record (EHR) and IBM MarketscanHealth Commercial Claims and Encounters (CCAE) administrative claims database. Data were standardized using the OHDSI Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

After characterization, we used the PLP package to train an L1-regularized logistic regression model to

[1] http://atlas-demo.ohdsi.org/#/cohortdefinition/1777320/, http://atlas-demo.ohdsi.org/#/cohortdefinition/1777321/, http://atlas-demo.ohdsi.org/#/cohortdefinition/1777322/

predict HF-related hospitalization among our target cohorts. We focused on the outcome of HF-related hospitalization among two target cohorts: TZD users with and without prior HF instances. Based on literature on TZD effects, we established the time-at-risk for HF-related hospitalization to be between 60 and 730 days after starting TZDs. Covariates included gender, age, index month, condition, drug group, procedure, and measurement time-bound covariates. We evaluated our model based on area under the ROC curve (AUC) and inspection of top-weighted features.

Results

Table 1. Demographics of patients by data source.

	CUIMC (N=1,330)		CCAE (N=29,031)	
	History of HF (N=207)	No History of HF (N=1,123)	History of HF (N=924)	No History of HF (N=28,107)
Age, years, median	66	62	59	56
Women, %	45.89	46.48	35.17	36.24
Time in cohort, days	261.96	348.97	164.01	182.69

Our phenotype identified a total of 30,361 patients with a history of HF (Table 1). Top drug occurrences for patients with a history of HF were aspirin, hydrochlorothiazide, and furosemide; top drug occurrences among patients without a history of HF were metformin and glucose (Table 2). Diabetes mellitus and hypertension were among top condition occurrences for both groups, though hypertension was more prevalent among patients with history of HF. Based on the initial ATLAS characterization, our TZD target cohorts appeared to differentiate patients with and without HF. Patients without HF did not appear to have medications or conditions to indicate otherwise, suggesting face validity.

Table 2. Top drug and condition occurrences of patients by data source.

	CUIMC (N=1,330)		CCAE (N=29,031)	
	History of HF (N=207)	No History of HF (N=1,123)	History of HF (N=924)	No History of HF (N=28,107)
Top five drug occurrences, descending order	Acetaminophen 325 MG (53.62%)	Acetaminophen 325 MG (32.15%)	Metformin Hydrochloride 1000 MG (33.23%)	Metformin Hydrochloride 1000 MG (34.41%)
	Docusate Sodium 100 MG (40.10%)	Metformin Hydrochloride 1000 MG (29.65%)	6 pack (Azithromycin 250 MG Oral Tablet (32.03%)	Pioglitazone 30 MG Tablet (29.53%)
	Aspirin 325 MG Delayed Release (32.85%)	Metformin Hydrochloride 500 MG (27.16%)	Influenza, seasonal, injectable (30.95%)	Pioglitazone 30 MG Tablet (29.00%)
	Acetaminophen 325 MG / Oxycodone Hydrochlorothiazide (30.43%)	Glucose 0.4 MG Oral Gel (24.13%)	Furosemide 40 MG Oral Tablet (30.19%)	Pioglitazone 30 MG Tablet (28.60%)
	Hydrochlorothiazide 25 MG (29.47%)	Pioglitazone 15 MG Tablet (23.33%)	Metformin Hydrochloride 500 MG (28.79%)	Metformin Hydrochloride 500 MG (28.00%)
Top five condition	Type 2 Diabetes without Complication (98.55%)	Type 2 Diabetes Mellitus w/o Complication	Diabetes mellitus without complication (92.86%)	Diabetes Mellitus w/o Complication (81.13%)

occurrences, descending order		(97.51%)		
	Essential Hypertension (91.79%)	Essential Hypertension (81.48%)	Type 2 diabetes mellitus (84.85%)	Type 2 Diabetes Mellitus (74.58%)
	Congestive Heart Failure (75.85%)	Type 2 Diabetes without Complication (66.87%)	Essential hypertension (83.77%)	Type 2 Diabetes Mellitus (67.11%)
	Essential Hypertension (74.40%)	Type 2 Diabetes without Complication (66.70%)	Essential hypertension (78.68%)	Type 2 Diabetes Mellitus (67.11%)
	Essential Hypertension (74.40%)	Essential Hypertension (63.86%)	Essential hypertension (78.68%)	Essential Hypertension (66.95%)

Our PLP model for predicting HF-related hospitalization among patients with history of HF produced an AUC of 0.55 and 0.756 on CUIMC and CCAE datasets, respectively. The PLP model for patients without history of HF resulted in AUCs of 0.527 and 0.804 on CUIMC and CCAE datasets, respectively (Table 3). For patients without prior HF, history of heart disease (β = 0.411) and use of beta-blockers (β = 0.345) were significant predictors of HF-related hospitalization in the model using CUIMC data. Similarly, furosemide, a diuretic frequently used to treat HF (β = 0.683), and beta-blockers (β = 0.427) were top predictors in the model generated using CCAE data.

Table 3. Patient Level Prediction Output

	CUIMC (N=1,330)		**CCAE (N=29,031)**	
	History of HF (N=207)	**No History of HF (N=1,123)**	**History of HF (N=924)**	**No History of HF (N=28,107)**
Accuracy (AUC)	0.55	0.527	0.756	0.804
Top Level Covariates, **beta**	drug_era: Topical treatment for hemmoroids or anal fissures **(0.6414)** condition_era: Heart failure **(0.2908)** drug_era: Vasoprotectives **(0.2459)** condition_era: Congestive heart failure **(0.2326)** ---	condition_era: Heart disease **(0.4112)** drug_era: Beta blocking agents, selective, and other diuretics **(0.3446)** drug_era: Blood and Blood Forming Organs **(0.2557)** gender = Male **(0.2318)** condition_era: Measurement finding above reference range **(-0.0933)**	procedure_occurrence: Initial hospital care, evaluation and management **(0.47)** condition_era: Neuropathy **(0.2285)** drug_era: Furosemide **(0.2587)** condition_era: Chronic heart failure **(0.2066)** drug_era: Diuretics **(0.2053)**	drug_era: Furosemide **(0.6826)** condition_era: Ulcer of lower extremity **(0.6579)** condition_era: Dyspnea **(0.4819)** drug_era: Beta Blocking Agents **(0.4274)** drug_era: Nervous System **(0.421)**

The presence of HF-related medications among top-weighted features in our PLP model for patients without history of HF suggests that re-evaluation of our original HF phenotype construction is merited. In particular, the presence of HF-related codes as features indicates the potential for error due to index date misclassification (*i.e.*, patients receiving treatment for HF-related conditions prior to the original index date were erroneously classified as having no history of HF). Additionally, both patient characterization and PLP suggested that some of the broad codes observed prior to the index date may indicate prior heart failure. Phenotypes developed specifically to identify first onset of chronic, progressive conditions such as

HF may be particularly prone to these issues.

Conclusions

This work underscores the importance of continuous phenotype evaluation after initial characterization for robust cohort definition. We offer several recommendations to support future phenotype development. First, incorporate and test multiple time windows when establishing a look-back period for phenotype definitions if calculating incidence rates. Second, leverage existing OHDSI tools such as Cohort Diagnostics to understand potential inconsistencies, particularly ones that are temporal in nature. Third, consider utilizing the PLP package as a downstream check of outcome and target phenotype definitions by reviewing features that appear most predictive in the models produced.

References

1. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc. 2013;20(1):117-21. PMCID: PMC3555337
2. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, Lai AM. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc. 2014 Mar-Apr;21(2):221-30. PMCID: PMC3932460
3. Hripcsak G, Albers DJ. High-fidelity phenotyping: richness and freedom from bias. J Am Med Inform Assoc. 2018 Mar 1;25(3):289-294. doi: 10.1093/jamia/ocx110. PMID: 29040596; PMCID: PMC7282504.
4. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, Basford M, Chute CG, Kullo IJ, Li R, Pacheco JA, Rasmussen LV, Spangler L, Denny JC. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc. 2013 Jun;20(e1):e147-54. PMID: 23531748
5. Nesto RW, Bell D, Bonow RO, Fonseca V, Grundy SM, Horton ES, et al. Thiazolidinedione Use, Fluid Retention, and Congestive Heart Failure: A consensus statement from the American Heart Association and American Diabetes Association. Diabetes Care. 2004 Jan 1;27(1):256–63.