Evaluating Phenotype Algorithms using PheValuator

Case Definitions and Phenotyping Algorithms

- "A case definition describes characteristics that a patient must possess to have a disease from a clinical perspective."

- "An EHR phenotyping algorithm is the translation of the case definition into an executable algorithm that involves querying clinical data elements from the EHR."

A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury

Casey Lynnette Overby,[1,2] Jon Duke,[3] Chi Gottesman,[4,5] Kay Wise,[1] Adler Perotte,[1] Chien Shing Ma,[3] Kevin Bruce,[3] Stephanie Johnson,[6] Jayant Talwalkar,[6] Yufeng Shen,[1,7] Steve Ellis,[5,8] Iftikhar Kullo,[6] Christopher Chute,[6] Carol Friedman,[1] Erwin Bottinger,[5,9,10] George Hripcsak,[1] and Chunhua Weng[1]

What is a phenotype algorithm and why do we need them

- Tendency to equate the case definition with the phenotype algorithm (or the cohort definition) – the algorithm is the coded *approximation* of the case definition.

- Case definitions must be translated into algorithms for working with observational datasets

- There can be loss in translation in creating a phenotype algorithm from a case definition

- How much inaccuracy? → Need for validation

Validating a Phenotype Algorithm

		Case Definition	
		Case	Non-Case
Phenotype Algorithm	Included	True Positive (TP)	False Positive (FP)
	Not Included	False Negative (FN)	True Negative (TN)

Ex.: True Positive (TP) – when a subject included in the phenotype algorithm is a case

For a complete validation of the algorithm we need:
1) Sensitivity: TP / (TP + FN)
2) Specificity: TN / (TN + FP)
3) Positive Predictive Value (PPV): TP / (TP + FP)
4) Negative Predictive Value (NPV): TN / (TN + FN)

Validating Algorithms

- Many research studies have attempted to validate algorithms
- Traditional validation involves chart reviews of patients by clinical experts
  - Time consuming
  - Costly
  - At the end, generally only determine PPV
- Needed a replacement that could do this quicker, easier, and produce all the elements of validation (i.e., sensitivity, specificity, PPV)

PheValuator: Development and evaluation of a phenotype algorithm evaluator

Joel N. Swerdel[a,b,*], George Hripcsak[b,c], Patrick B. Ryan[a,b,c]

[a] Janssen Research & Development, 920 Route 202, Raritan, NJ 08869, USA
[b] OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), 622 West 168th Street, PH-20, New York, NY 10032, USA
[c] Columbia University, 622 West 168th Street, PH20, New York, NY 10032, USA

ARTICLE INFO

ABSTRACT

Background: The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. However, a complete evaluation of PAs, i.e., determining sensitivity, specificity, and positive predictive value (PPV), is rarely performed. In this study, we propose a tool (PheValuator) to efficiently estimate a complete PA evaluation.

Methods: We used 4 administrative claims datasets: OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN); IBM MarketScan Multi-State Medicaid); IBM MarketScan Medicare Supplemental Beneficiaries; and IBM MarketScan Commercial Claims and Encounters from 2000 to 2017. Using PheValuator involves (1) creating a diagnostic predictive model for the phenotype, (2) applying the model to a large set of randomly selected subjects, and (3) comparing each subject's predicted probability for the phenotype to inclusion/exclusion in PAs. We used the predictions as a 'probabilistic gold standard' measure to classify positive/negative cases. We examined 4 phenotypes: myocardial infarction, cerebral infarction, chronic kidney disease, and atrial fibrillation. We examined several PAs for each phenotype including 1-time (1X) occurrence of the diagnosis code in the subject's record and 1-time occurrence of the diagnosis in an inpatient setting with the diagnosis code as the primary reason for admission (1X-IP-1stPos).

Results: Across phenotypes, the 1X PA showed the highest sensitivity/lowest PPV among all PAs. 1X-IP-1stPos yielded the highest PPV/lowest sensitivity. Specificity was very high across algorithms. We found similar results between algorithms across datasets.

Conclusion: PheValuator appears to show promise as a tool to estimate PA performance characteristics.

Overview of PheValuator Process
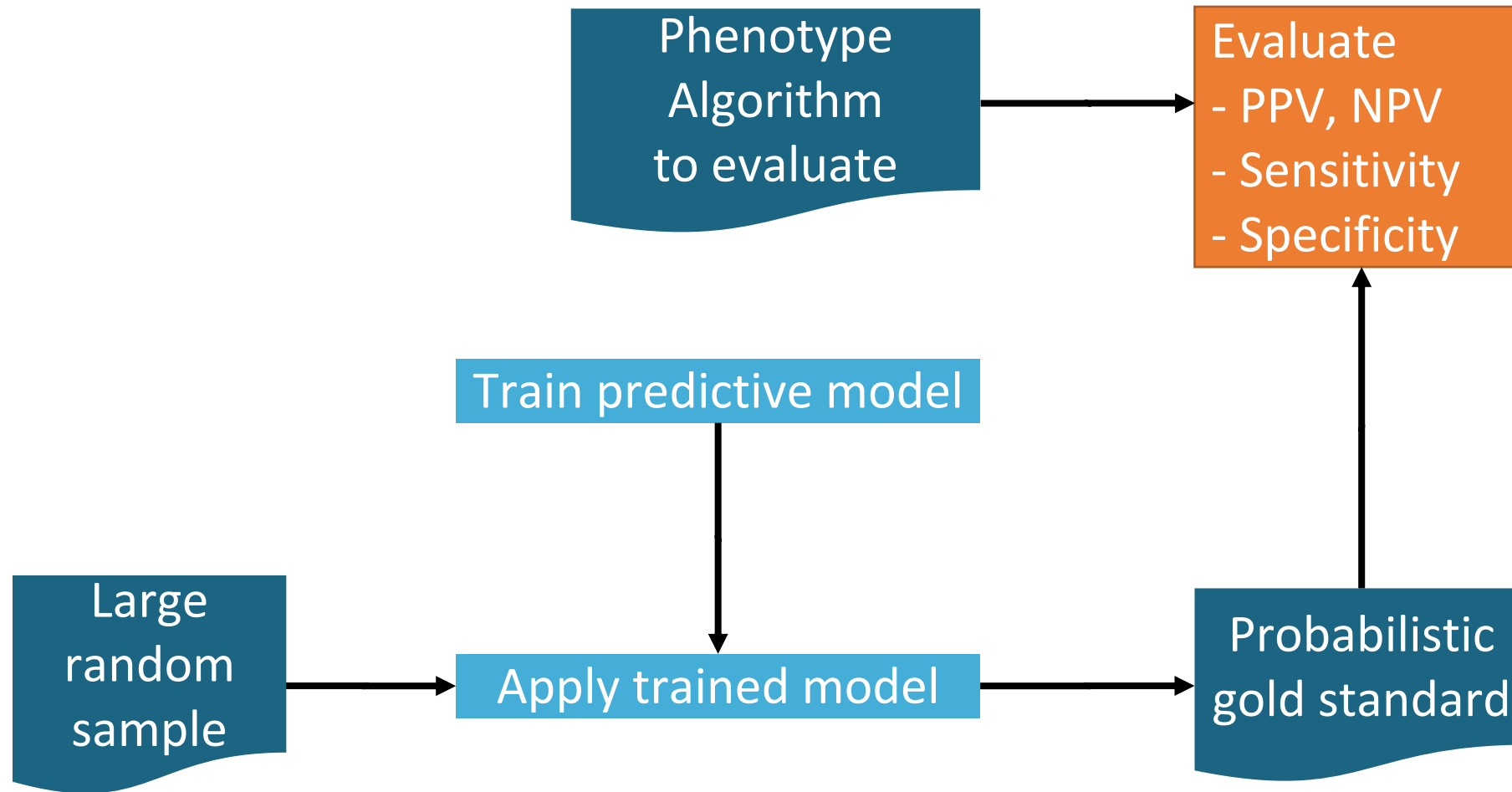
Overview of PheValuator

1. Create a subject population of cases and non-cases for the model
2. Extract health data from the overall population – both cases and non-cases
3. Use regularized logistic regression (LASSO) to develop a model to discriminate between cases and non-cases
4. Use the model to determine the probability of subjects having the health outcome
5. Evaluate the phenotype algorithm

Example Results

Downloading PheValuator

https://github.com/OHDSI/PheValuator

Limitations of Phenotype Evaluation

- It's only as good as the data
  - PheValuator estimates a probability of a health outcome similar to a clinician reviewing the data – if the data is sparse, there is less information to assess
- Some diseases are poorly differentiated from other diseases
  - E.g., Similar symptoms, treatment
  - Diagnostic testing is weak or non-existent

Where PheValuator has been Applied

In Public Domain
- PheValuator Journal Article
  - Myocardial Infarction
  - Ischemic Stroke
  - Atrial Fibrillation
  - Chronic Kidney Disease
- ICPE
  - Pulmonary Arterial Hypertension
  - Chronic Thromboembolic Pulmonary Hypertension
- OHDSI Symposium
  - Acute Renal Failure
  - Kidney Stone
  - Renal Cell carcinoma
  - Atopic Dermatitis
  - Psoriasis
  - Candidiasis
  - Melanoma

Internal Use
- Rheumatoid Arthritis
- Anemia
- Ankylosing Spondylitis
- Malignancies
- Crohn's Disease
- Ulcerative Colitis
- GI Bleed
- Hemorrhagic Stroke
- Hyperlipidemia
- Hypothyroidism
- Depression
- Schizophrenia
- Serious Infection
- Blood Dyscrasias