

A Novel Phenotyping Pipeline To Improve Identification Of Patients With Pulmonary Hypertension In Electronic Health Records

D. M. Vidmar¹, W. Thompson¹, K. Morland², G. Lee¹, R. Miotto¹, J. Leader¹, J. Pfeifer¹, B. Fornwalt¹, A. Nelsen², R. Chen¹
¹Tempus AI, Inc., Chicago, IL, United States, ²United Therapeutics, Research Triangle Park, NC, United States

INTRODUCTION

- Accurately identifying patients with pulmonary hypertension (PH) is critical for defining clinical research cohorts and generating accurate labels for machine learning (ML) models.
- PH is often poorly captured in observational datasets due to known limitations of structured data such as diagnosis codes or problem lists.
- We hypothesized that a phenotype built on structured and unstructured electronic health record (EHR) data could outperform a comparable structured phenotype baseline.

METHODS

- Using a longitudinal de-identified EHR dataset of over 2 million patients, we first created and iteratively improved a structured phenotype using various data elements including diagnosis codes, problem lists, procedures, medications, and imaging results.
- We then developed labeling functions - a collection of logic rules and regular expressions to determine whether patients are likely to have PH. These labeling functions integrated multimodal signals from structured data, unstructured note data, and diagnostic study reports and were aggregated to generate weak labels.
- Then, we used these weak labels to train a state-of-the-art transformer-based natural language processing (NLP) model. All structured and ML phenotypes were iteratively developed on a training data set and then evaluated on an independent test set; both training and test sets underwent blinded, manual chart review by clinicians (total N=600).

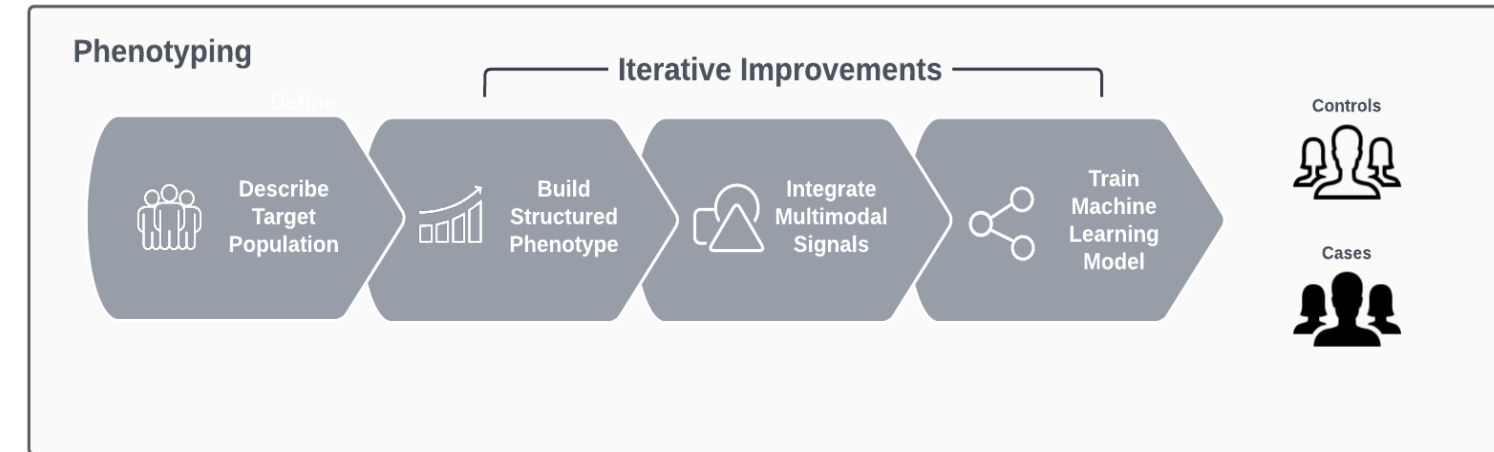


Figure 1. Phenotyping pipeline process diagram.

SUMMARY

- Identifying patients diagnosed with pulmonary hypertension in EHR data is challenging due to non-specific diagnosis codes
- We propose a pipeline that applies and aggregates 21 different heuristic rules across various EHR data elements
- The resulting phenotype outperforms a structured phenotype on independent holdout data

RESULTS

I. Patient Extraction Stage

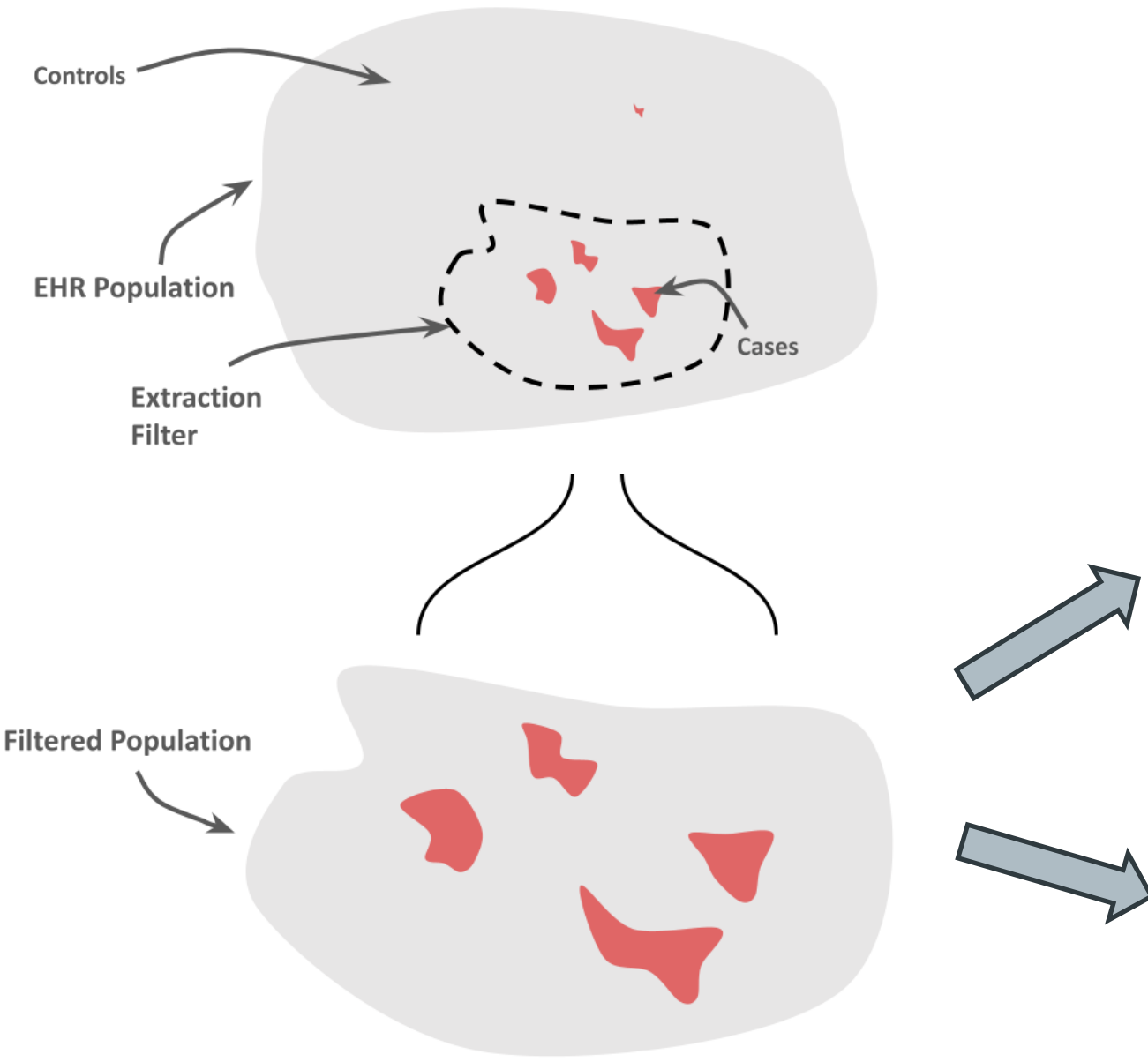


Figure 2. We enrich our cohort by extracting all patients who have at least one piece of evidence suggesting a possible PH diagnosis, spanning broadly-defined text searches (for example "pulmonary hypertension", "pulm htn", "pHTN"), diagnosis codes, and medications. We assume that an insignificant number of patients with a PH diagnosis are missed at this stage, with patients passing this filter being considered potential candidates.

II. Structured Phenotype Baseline

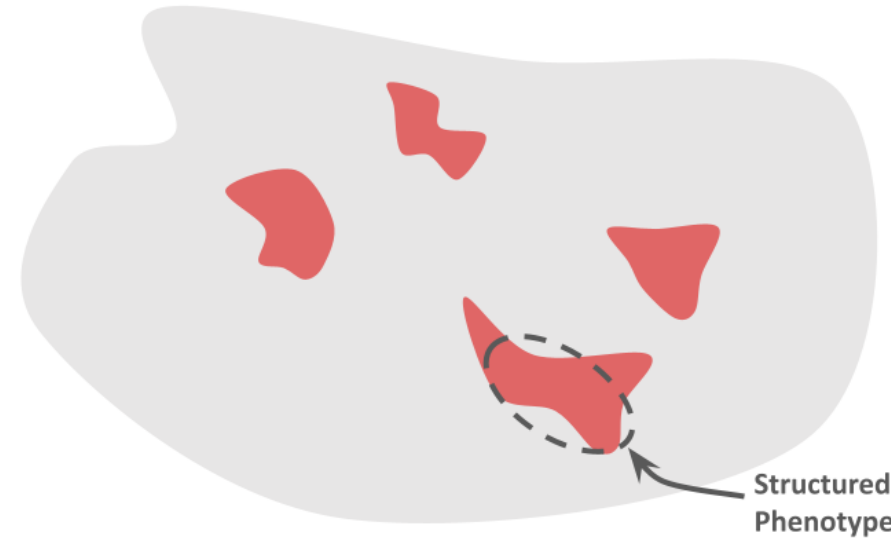


Figure 3. Structured phenotypes are comprised of a single set of rigid rules (for example a set of diagnosis codes). This can make it difficult to capture complex boundaries between cases and controls.

Phenotype Performance Comparison

	PPV	Sensitivity	F1 score
Structured Phenotype	0.93	0.54	0.68
Weak Labels	0.84	0.74	0.78
NLP Neural Net Model	0.93	0.70	0.80
NLP Model + Structured	0.91	0.76	0.83

Table 1. Performance metrics on the holdout set. Phenotype performance is judged as a balance between PPV and sensitivity, with higher F1 scores representing better overall performance.

III. Weak Labeling Pipeline

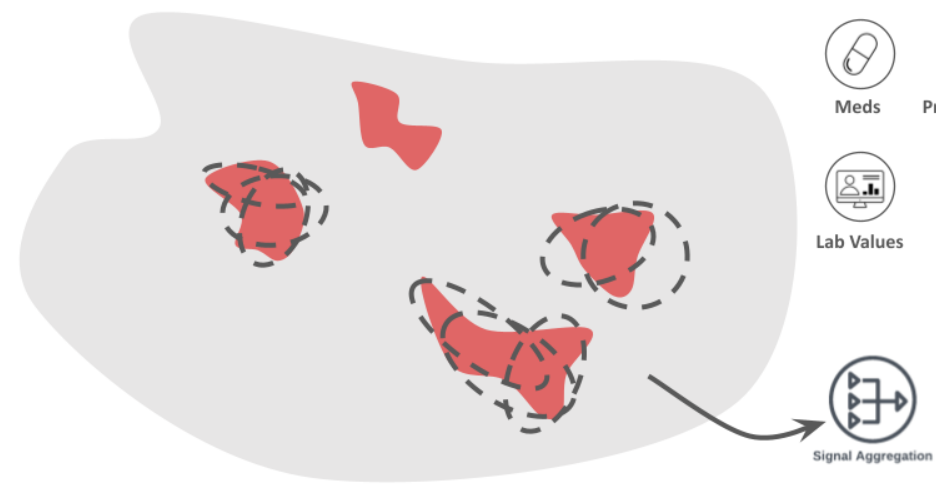


Figure 4A. We apply our multimodal labeling functions to filtered data. The resulting votes are aggregated to a probabilistic weak label for each patient.

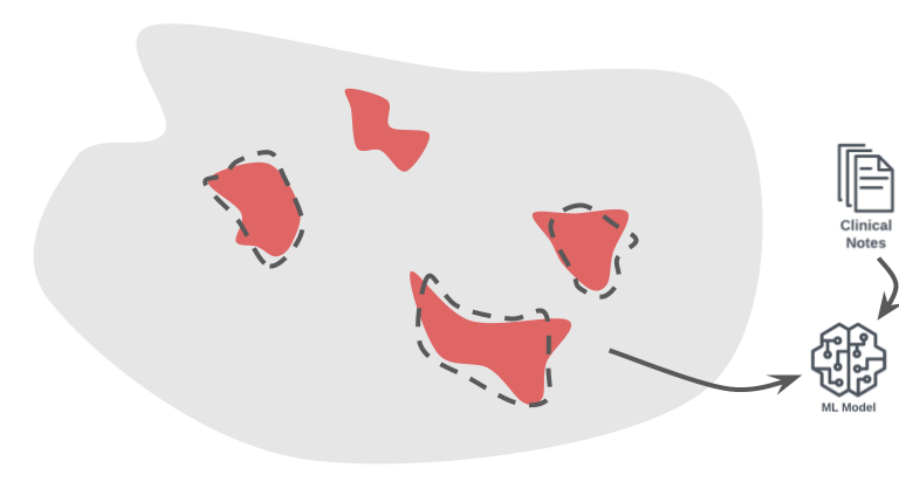


Figure 4B. We train an NLP model across clinical note text for each patient. We use the weak labels from the previous step in our loss function.

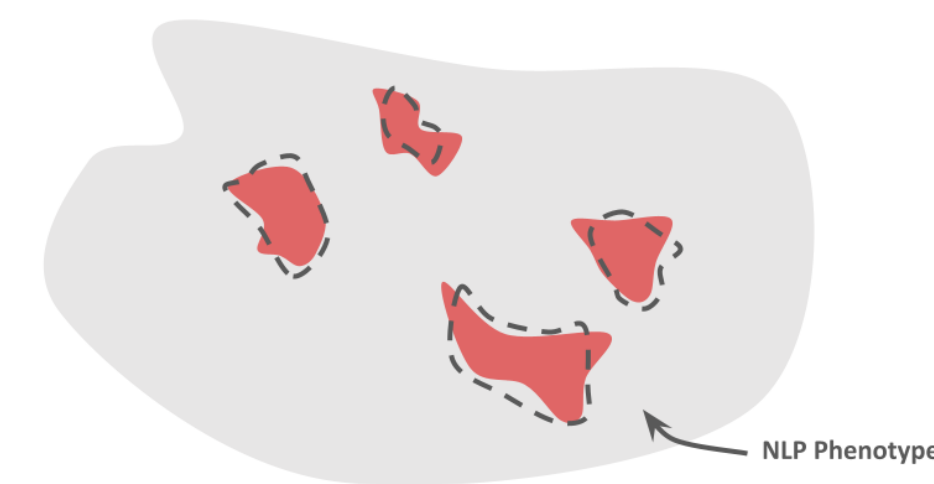


Figure 4C. We predict patient labels using this model to get our final phenotype. Since this model is regularized and analyzes all note text, it can outperform the weak labels.

Conference specific guidelines

Guidelines from 2023

- **No excessive use of jargon or abbreviation**
- **No brand or company logos**

Internal due dates:

- Posters ready for SciComm review: 4/26
- Posters submitted for legal/leadership review: 5/03
- Conference: 5/17

Data Visualization Guidelines

Tempus Color palettes

Qualitative

SciComm preferred palette

This is a minor update to the default palette recommended by graphic design (see below) chosen to minimize the grouping of similar colors.



[#5993F7', '#D97C4F', '#62B882', '#CC78A7', '#774D9A', '#515CBE', '#E9C74E', '#B8E382', '#A54A72', '#C8B1F6']

Graphic design recommendation

For cases where data are paired or grouped in a logical way, we recommend using this ordering (or any re-ordering) that results in the clearest presentation of the data



[#5993F7', '#515CBE', '#D97C4F', '#E9C74E', '#774D9A', '#C8B1F6', '#A54A72', '#CC78A7', '#62B882', '#B8E382']

Graphic design variant

In the event that a slightly lighter look is preferred, this palette (or a logical re-ordering of colors to fit the application) is acceptable



[#86B2FF', '#738AFF', '#F99B6D', '#FCE285', '#AD6CE4', '#CCB2FF', '#E777A8', '#FFC0E3', '#89D3A5', '#D1ECAFF']

Continuous

(Note: while these palettes are meant to be used in continuous applications, they are ultimately constructed from discrete color palettes with code examples showing how to properly extrapolate and create a continuous palette for applications such as heatmaps. However, these palettes may also be used in their discrete form [depending on the application], much the same as the qualitative palettes listed above.)

Sequential



[#29293C', '#384162', '#485889', '#5770AF', '#6687D6', '#779BEB', '#8BACED', '#9EBDF0', '#B3CCF3', '#C7DDF6']

Diverging



[#384162', '#475889', '#5770AF', '#6687D6', '#8BACED', '#FFFFFF', '#DDC2CD', '#C99EAD', '#B47A8F', '#9F5773', '#893157']

- [SciComms Data Visualization Best Practices](#)
- Please use Roboto font
- [Figure Sizing and Exporting](#)