# MINING FOR HEART FAILURE SYMPTOMS IN CLINICAL NOTES

Sadiq G[1], Oguntuga A[2], Sudaria T[2], Hubbard A[1], Overcash J[1], Nguyen N[2]

[1] Veradigm® Raleigh, NC, USA [2] Veradigm™ San Francisco, CA, USA

## BACKGROUND

Heart failure (HF) is a chronic disease that often worsens over time. Signs and symptoms are of diagnostic value for classifying severity of disease and recommending appropriate treatments. Symptoms are often lacking in structured fields in electronic health record (EHR) platforms but may be recorded in unstructured clinical notes, also known as SOAP (Subjective, Objective, Assessment, and Plan) notes.[1] Extracted symptoms may aid researchers in the determination of severity or stage of heart failure, when sufficient information for doing so is not available in structured fields alone.

## OBJECTIVES

We extracted HF symptoms from clinical notes in a large cloud-based EHR for ambulatory patient care offered by Veradigm using Natural Language Processing (NLP) methods to enrich structured information already available in the EHR database.

## METHODS

Twenty-four (24) HF signs and symptoms[2,3] were identified with support from clinical subject matter experts.

- Cough
- Dyspnea
- Fatigue
- Palpitation
- Sleep apnea
- Presyncope
- Hepatomegaly
- ICD shock
- Ventricular tachycardia
- Jugular venous pressure
- Heart murmur
- Anorexia
- Cardiac cachexia
- Natriuretic peptide biomarkers
- Ankle edema
- Leg edema
- Peripheral edema
- Pulmonary edema
- Exercise intolerance
- Chest pain
- Family history of cardiomyopathy
- Right ventricular heave
- Cold lower extremities
- Taking extended release metoprolol succinate

To evaluate how each HF symptom is used in the EHR, 1,433 sentences were manually reviewed. Synonyms, abbreviations, singular, and plural words were noted, and regular expressions were created as offsets for each symptom.

For example, synonyms for “natriuretic peptide biomarkers” included:

- Natriuretic Peptide Biomarkers
- NPs
- brain natriuretic peptide
- b-natriuretic peptide
- B-type natriuretic peptide
- B type natriuretic peptide
- Natriuretic Peptide
- bnp
- nt-probnp

After manual review and curation of regular expressions, the following steps were taken:

- Physician notes in the EHR were short-listed for those containing at least one HF symptom.
- Shortlisted notes were further broken into sentences based on punctuation: new-line characters, periods, exclamation points, semi-colons, and question marks.
- Sentences were lemmatized by sorting words grouped by inflected or variant forms of the same word using the Python package, Spacy.[4]
- Symptoms and their offsets were found and extracted using Regular Expressions.
- Negation of symptoms were checked with a rule-based algorithm.
- Symptoms were mapped to SNOMED, ICD9, and ICD10 codes and organized to structured query language (SQL) tables.
- Algorithms are rerun once a month for regular extracts starting 1 January 2020.

## LIMITATIONS

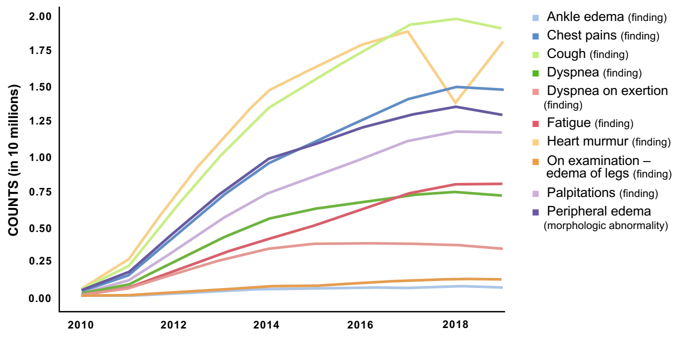
**This study is subject to the following limitations:**

- The Veradigm EHR platform examined in this study is used primarily in small, independent practices, therefore results may not be generalizable to other care settings.

## RESULTS

There were 35M patients with over 2B physician notes in the EHR; approximately 15% of these notes contained a potential HF symptom. In structured data, 8M instances of HF symptoms were recorded. When notes were enriched using NLP, another 130M instances of these symptoms were discovered (a 16-fold increase). The biggest gains were in heart murmur, sleep apnea, and right ventricular heave. Certain symptoms (e.g., cardiomyopathy family history and jugular venous pressure) were found only in clinical notes.

**Figure 1 | Top 10 most common HF symptoms per year in an ambulatory practice EHR**




There were 883.7K patients with ICD-9-CM diagnosis codes for HF. In structured data they had only 512K recorded symptoms. Clinical notes had 9.3M additional recorded symptoms, an increase of 18-fold. The rules-based NLP method used in this study yielded Accuracy=96%, Recall=97%, Precision=98.5%, and F1 Score=98%.

**Figure 2 | Number of Transcripts with HF Symptoms per Year**

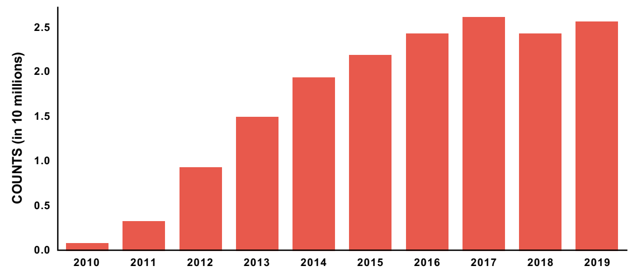
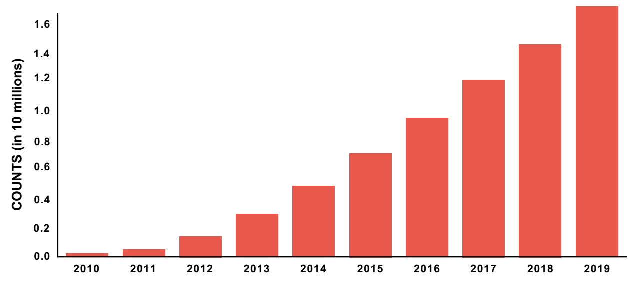



**Figure 3 | Cumulative Number of Transcripts with HF Symptoms per Year**




## CONCLUSION

Signs and symptoms that are entered into notes are critical to patient assessment and development of guideline-based treatment plans. Other disease symptoms might also be mined with similar methods. HF signs and symptoms were successfully extracted from unstructured clinical notes using a rules-based algorithm

- The count of symptoms recovered from notes exceeded the data in structured fields by more than 15-fold.

In the future, the performance of machine learning-based algorithms to extract disease symptoms may be compared with the performance of rules-based algorithms.

## REFERENCES

1. Meystre SM, Kim Y, Gobbel GT, Matheny ME, Redd A, Bray BE, Garvin JH. Congestive heart failure information extraction framework for automated treatment performance measures assessment. J Am Med Inform Assoc. 2017 Apr 1;24(e1):e40–6. doi: 10.1093/jamia/ocw097.
2. Vijayakrishnan R, Steinhubl SR, Ng K, et al. Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. J Card Fail. 2014;20(7):459-464. doi:10.1016/j.cardfail.2014.03.008
3. Watson RD, Gibbs CR, Lip GY. ABC of heart failure. Clinical features and complications. BMJ. 2000;320(7229):236-239. doi:10.1136/bmj.320.7229.236
4. https://spacy.io/

**International Society for Pharmacoeconomics and Outcomes Research (Virtual) ; Bay Lake, FL, USA; May 18 – 20, 2020**