

Using Natural Language Processing to Facilitate Medical Record Abstraction in Epidemiological Studies.

The Atherosclerosis Risk in Communities (ARIC) Study

C. Moore¹, K. Shaffer¹, N. Puccinelli-Ortega², A. Kucharska-Newton¹, S. Haas¹, G. Heiss¹.

University of North Carolina at Chapel Hill¹, Wake Forest School of Medicine², NC



INTRODUCTION

- The Atherosclerosis Risk in Communities (ARIC) study conducts ongoing surveillance of hospitalized cardiovascular events and death in 4 communities in the United States (NC, MS, MN and MD).
- Diagnostic criteria for heart failure (HF), including the presence of symptoms consistent with HF decompensation (e.g., worsening shortness of breath) during hospitalizations, are manually abstracted from medical records.
- Manual chart abstraction occurs under a stringent protocol and has high repeatability, but it is time consuming and costly.
- Study Objective.** Develop and test natural language processing (NLP) tools that extract information on symptoms of HF from free-text electronic medical records.

METHODS

- We used c-TAKES (clinical Text Analysis and Knowledge Extraction Tool), open-source NLP software, to identify the presence of HF symptoms of worsening shortness of breath, edema, paroxysmal nocturnal dyspnea, and orthopnea in de-identified free-text electronic medical records from the ARIC study.
- We then created a program, written in Python, that processes the c-TAKES output to negate relevant concepts not consistent with worsening symptoms. For example, the shortness of breath concept in the text “the patient has chronic shortness of breath” is negated because the symptom is not worsening.
- Using an initial corpus of 112 medical records (representing 112 unique hospitalizations) adjudicated by ARIC chart abstractors as the “gold” standard, we evaluated recall and precision of the NLP tool.
- An independent reviewer investigated instances of discordance between the “gold” standard assessment and NLP results.

RESULTS

TABLE 1. NLP Performance Characteristics (n=112)

ARIC HF Variable	Recall	Precision	# of patients that symptoms identified by NLP were overlooked by abstractors
Shortness of breath	98%	76% (91%) [†]	14
Edema	98%	53% (66%) [†]	10
Paroxysmal nocturnal dyspnea	86%	64% (73%) [†]	2
Orthopnea	82%	81% (90%) [†]	1

[†] corrected for instances in which symptoms identified by NLP were overlooked by abstractors

TABLE 2. Examples of ARIC Abstractor Missing Symptoms Identified By NLP

ARIC HF Variables	Relevant Free-Text
Shortness of breath	“Has had runs of V-tach overnight with shortness of breath. EKG overnight was unchanged. Troponins continue to be trended.”
Edema	“Her sx are c/w volume overload, severe valvular disease. Bilateral lower extremity edema 2+”
Paroxysmal nocturnal dyspnea	“Frequent PND and pedal edema”
Orthopnea	“Does endorse some orthopnea”

TABLE 3. Examples of NLP Missing Symptoms Identified by Abstractors

ARIC HF Variables	Relevant Free-Text
Shortness of breath	“Resp: [15-29] 20”
Paroxysmal nocturnal dyspnea	“He has been unable to sleep at night and frequently wakes from sleep short of breath.”
Orthopnea	<ul style="list-style-type: none">“The symptoms are aggravated by activity and a supine position.”“Shortness of breath is worse when lying flat.”“He states that he cannot lay down flat due to shortness of breath.”

RESULTS SUMMARY

- Table 1 shows recall and precision of NLP vs. ARIC abstractors in identifying HF symptoms.
- The NLP program performed well (Tables 1 and 2) in identifying HF symptoms within the free-text of physician notes and reports. Those notes are read by ARIC chart abstractors who may occasionally miss occurrence of HF symptoms.
 - for example, there were 14 instances in which NLP identified shortness of breath symptoms missed by abstractors (Table 1).
- Areas in which NLP missed HF symptoms identified by abstractors (Table 3) relate to instances in which:
 - symptoms are described, but not explicitly stated (e.g., “shortness of breath while lying flat” describing orthopnea)
 - symptoms are implied based on abnormal vital signs (e.g., “respiratory rate > 22 breaths/min” implying shortness of breath)

DISCUSSION

- We developed an NLP tool that has good recall and precision for identifying HF symptoms in free-text electronic medical records.
- The NLP tool is being refined to further improve its performance characteristics and to extract additional ARIC HF symptoms embedded in free-text medical records.
- Our results suggest that use of validated NLP technology holds the potential for improving the cost-effectiveness of epidemiologic surveillance studies.

ACKNOWLEDGEMENT

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts: HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C.

The authors thank the staff and participants of the ARIC study for their important contributions