

# A Linguistic Analysis Examining the Impact of COVID-19 on Pneumonia Diagnosis and Disease Models

Alec B CHAPMAN<sup>a,b,1</sup>, Kelly S PETERSON<sup>a,b</sup>, Elizabeth RUTTER<sup>a</sup>, Mckenna NEVERS<sup>a,b</sup>, Jian YING<sup>a,b</sup>, David CLASSEN<sup>a,b</sup>, Makoto JONES<sup>a,b</sup>, Matthew SAMORE<sup>a,b</sup> and Barbara JONES<sup>a,b</sup>

<sup>a</sup>*Veterans Affairs (VA) Salt Lake City Health Care System, Salt Lake City, UT, US*

<sup>b</sup>*University of Utah School of Medicine, Salt Lake City, UT, US*

**Abstract.** Written clinical language embodies and reflects the clinician's mental models of disease. Prior to the COVID-19 pandemic, pneumonia was shifting away from concern for healthcare-associated pneumonia and toward recognition of heterogeneity of pathogens and host response. How these models are reflected in clinical language or whether they were impacted by the pandemic has not been studied. We aimed to assess changes in the language used to describe pneumonia following the COVID-19 pandemic.

**Keywords.** Biosurveillance, COVID-19, natural language processing and text mining, population health

## 1. Introduction

Written clinical language embodies and reflects the clinician's mental models of disease. Prior to the COVID-19 pandemic, pneumonia was shifting away from concern for healthcare-associated pneumonia ("HCAP") and toward recognition of heterogeneity of pathogens and host response. How these models are reflected in clinical language or whether they were impacted by the pandemic has not been studied. We aimed to (1) examine the impact of the pandemic on pneumonia diagnosis using natural language processing (NLP); and (2) assess changes in language using linguistic analysis of clinical text.

## 2. Methods

We identified all hospitalizations in the US Department of Veterans Affairs (VA) from the Emergency Department (ED) occurring 1/1/2015-12/31/21 with a diagnosis of pneumonia using a previously validated approach that combines diagnosis codes with NLP-extracted diagnostic assertions of pneumonia from clinical notes and chest imaging reports [1]. We examined trends in the prevalence of pneumonia diagnoses and diagnostic accuracy (positive predictive value [PPV] of clinical diagnosis against chest

---

<sup>1</sup> Corresponding Author: Alec Chapman, email: [alec.chapman@hsc.utah.edu](mailto:alec.chapman@hsc.utah.edu).

imaging). To evaluate pre- and post-pandemic trends in language, we conducted a linguistic analysis of the percentage of pneumonia positive documents.

### 3. Results

Among 2.4M hospitalizations, 363,324 (15.4%) had a clinical diagnosis of pneumonia. PPV against chest imaging was stable then increased slightly during the pandemic. Prevalence of pneumonia diagnoses varied seasonally but was stable across the 5 years prior to the pandemic (range= 14%-14.7%) then increased (17.9% in both 2020 and 2021). Over half of all documents contained qualifying terms of uncertainty; this increased, then plateaued during the pandemic. Temporal trends of phrases used to describe pneumonia are shown in Figure 1. The use of subtypes (i.e., “bacterial” and “viral”) steadily increased then accelerated during the pandemic. Use of “sepsis” increased prior to the pandemic and then decreased sharply. Other host-response qualifiers (i.e., “inflammation” and “ARDS”) were rarely used in 2015-2019 then increased during the pandemic. “HCAP” decreased dramatically during the study period and independent of the pandemic.

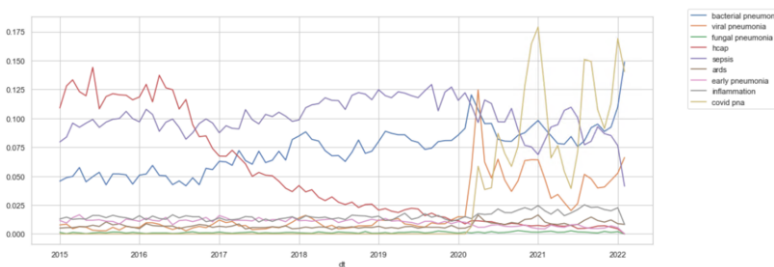


Figure 1. Monthly proportion of pneumonia-positive documents containing specific phrases and qualifiers.

### 4. Conclusions

Pneumonia prevalence was stable prior to and then increased during the pandemic. Language used to describe pneumonia diagnosis in clinical notes demonstrated a decline in HCAP, increased uncertainty, distinguishing bacterial from viral pneumonia, and increased characterizations of the host inflammatory response. Clinical text can reveal shifting mental models of disease, which may explain changes and variation in clinical behavior. Future work is needed to understand how language and behavior influence each other in different settings and whether the pandemic will have lasting impacts on our models of respiratory infection.

### References

- [1] Chapman AB, Peterson KS, Rutter E, Nevers M, Zhang M, Ying J, Jones M, Classen D, Jones B. Development and evaluation of an interoperable natural language processing system for identifying pneumonia across clinical settings of care and institutions. *JAMIA Open*. 2022 Dec;5(4):ooac114, doi: 10.1093/jamiaopen/ooac114.