

## **ONLINE SUPPLEMENT**

**Title:** Development of a Post-Acute Sequelae of COVID-19 (PASC) Symptom Lexicon Using Electronic Health Record Clinical Notes

**Authors:** Liqin Wang, PhD<sup>1,2</sup>; Dinah Foer, MD<sup>1,2,3</sup>; Erin MacPhaul, MS<sup>1</sup>; Ying-Chih Lo, MD<sup>1,2</sup>; David W. Bates, MD, MSc<sup>1,2</sup>; Li Zhou, MD, PhD<sup>1,2</sup>

### **Affiliations:**

<sup>1</sup> Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital

<sup>2</sup> Harvard Medical School, Boston, Massachusetts, USA

<sup>3</sup> Division of Allergy and Clinical Immunology, Department of Medicine, Brigham and Women's Hospital

### **Supplemental Material Contents:**

**eTable 1.** Challenges in a rule-based natural language processing (NLP) approach to symptom extraction from electronic health record (EHR) clinical notes

**eTable 1.** Challenges in a rule-based natural language processing (NLP) approach to symptom extraction from electronic health record (EHR) clinical notes

<b>EHR Feature</b>	<b>Example from clinical notes</b>	<b>Challenge for NLP</b>
Symptoms scores (Patient Reported Outcome Measures)	“Edinburgh Depression Scale score:2”	Requires medical knowledge to determine if the patient is having depression symptoms
Standard instructions	“If you find you often feel sad or hopeless, talk with your doctor”  “Symptoms of depression include: a sad or gloomy mood, lack of pleasure in activities you previously enjoyed, difficulties managing your weight, early morning awakening with inability to return to sleep or excessive sleeping, agitation or moving so slow that others notice, difficulty concentrating, fatigue, or mental slowness”	Often included in clinical notes as anticipatory guidance to patients based on reason for visit or medication prescriptions, regardless of current patient symptoms
Linguistic context	“Stress test”  “Stress fracture”  “The patient stressed that the nausea had not gone away despite dietary changes”	Symptom terminology may also be used in other medical contexts or as colloquialisms
Negation	“Mental Health Screening: Depression Concerns: none”  “Fever / Chills: n”  “(-) weight gain”	Negation terms may be distant from symptom term, or may be represented by abbreviations [n; (-)]
Past medical history	“DOE (dyspnea on exertion) 9/9/2019”	Symptoms preceded the study timeframe
Treatment indication	“albuterol 90 mcg/actuation inhaler Inhale 2 puffs into the lungs every 6 (six) hours as needed for wheezing or shortness of breath/dyspnea”	Medication instruction include clinical indications, which may be representative of active or anticipated patient symptoms
Allergic reaction	“Statins-hmg-coa reductase inhibitors Myalgia 10/06/2020”	Symptoms in the study timeframe but with a

		clear non-COVID-19 associated trigger
Review of systems	"[] Anorexia [] fatigue [x] insomnia [] fever [] chills [] dizziness [] weight gain [x] weight loss [] night sweats"	Clinical tabulation of current symptoms denoted with signs such as "+" or "x"
Implied symptoms	<p>"Encouraged to lose weight, portion control, hydration, and exercise will be key"</p> <p>"She has intentionally lost weight, reports an 8lb weight loss with cutting out soda and walking more"</p>	Therapeutic intervention implies presence of a symptom, e.g., a patient needs to lose weight, because they have had weight gain. Rule based NLP may have difficulty differentiating intentional and unintentional patient reported outcomes or symptoms