

**Topic Modeling of Patient Portal and Telephone Encounter Messages: Insights from a Cardiology Practice**

Jedrek Wosik, MD<sup>1</sup>, Shijing Si, PhD<sup>2</sup>, Ricardo Henao, PhD<sup>2</sup>, Mark Sendak, MD<sup>3</sup>, William Ratliff, MBA<sup>3</sup>, Suresh Balu, MBA<sup>3</sup>, Deepthi Krishnamaneni<sup>4</sup>, Ryan Craig<sup>4</sup>, Eric Poon, MD, MPH<sup>4</sup>, Lawrence Carin, PhD<sup>2</sup>, Manesh Patel, MD<sup>1</sup>

<sup>1</sup>Duke University Medical Center, Durham, NC; <sup>2</sup>Duke University, Durham, NC; <sup>3</sup>Duke Institute of Health Innovation, Durham, NC; <sup>4</sup>Duke Health Technology Solutions, Durham, NC

**Background.** Demand for non–face-to-face care, such as communication through a patient portal and telephone call, is increasing. Patient portals have been rapidly adopted across diverse clinical specialties; with some portal contacts exceeding traditional face-to-face outpatient encounters. With the introduction of patient portals, phone call volumes can increase and create more workload on the clinical systems. Understanding message content is needed to properly engage, route and respond to those messages. Natural language processing applied to this unstructured data can support patient care and reduce clinician workload.

**Methods.** This retrospective study was conducted at Duke University Health System (DUHS), a private, non-profit institution that provides primary and regional referral care for patients with cardiovascular (CV) disease or risk factors for CV disease, and over 140,000 yearly outpatient visits annually. A total of 167,030 adult patient portal messages and 822,597 telephone encounters messages were used in the analysis from 10/2014 – 08/2018 to their cardiology teams. These messages consisted of free, unstructured text. Using a Latent Dirichlet Allocation (LDA) topic model, we extracted 10 topics learned each across the full corpus of patient portal and telephone encounter messages. A UCI coherence score (C<sub>v</sub> of 0.64 and 0.57 for patient portal messages and telephone messages, respectively) was achieved with the optimal number of topics. The research team assigned topic interpretations based on clinician judgement.

**Results.**

Ten unique topics were identified for the patient portal messages:

PATIENT PORTAL MESSAGE TOPICS				
Administrative 22,125 (13.24%)	Afib/Afl 16,483 (9.86%)	Complex prescriptions 19,659 (11.77%)	Vitals 14,986 (8.97%)	Scheduling 20,327 (12.17%)
Medications and symptoms 15,539 (9.30%)	Simple prescription 17,495 (10.47%)	Clinical symptoms 15,061 (9.02%)	Results (labs/ procedures) 14,605 (8.74%)	Miscellaneous (social) 10,755 (6.43%)

Ten unique topics were identified for the telephone messages:

TELEPHONE MESSAGE TOPICS				
LVAD INR Note 50,425 (6.13%)	Scheduling, clinical follow-up 194,408 (23.63%)	Clinical questions, scheduling, appointments 147,731 (17.96%)	Questions about symptoms 66,876 (8.13%)	Discharge follow-up questions 22,982 (2.79%)
Lab results, interpretation/ recommendations 82,441 (10.02%)	Discharge follow-up questions 65,446 (7.96%)	Holter reports, Cardionet Summary 53,915 (6.55%)	CT surgery discharge follow-up, other questions 69,243 (8.42%)	Organ donation/Transplant Candidacy 69,130 (8.40%)

**Conclusion.**

Topic modeling of patient portal and telephone messages suggest that patients use these communication tools for different purposes. Although each messages can contain all 10 topics, there are distinct topics that can be used for tailored quality improvement interventions based on message topics and preferred method of communication. Focus on patient digital engagement while reducing clinician workloads can be targeted through self-service tools for scheduling, administrative tasks (i.e. insurance forms, referral requests) and medications refills. Improved telephone call triaging by topic as well as augmented agent response using natural language processing and artificial intelligence (i.e. monitor for emergency or trigger phrases, suggest responses) can further improve patient outcomes and health system efficiency.

