Why Patient Portal Messages Indicate Risk of Readmission for Patients with Ischemic Heart Disease

Lina Sulieman, PhD, Zhijun Yin, PhD, Bradley A. Malin, PhD
[1] Vanderbilt University, Nashville, TN

Abstract

Online portals enable patients to exchanging messages with healthcare providers. After discharge, patients message providers to ask questions and report problems. Care providers read and respond accordingly, which requires a non-trivial amount of human effort and is unlikely to scale up as portals become more popular. Automatically detecting when a message indicates a worsening in a patient's condition can assist providers to identify patients at risk of readmission. We investigated the association between messages that patients, diagnosed with ischemic heart disease, sent after discharge and the risk of readmission. We studied 4,052 messages sent after discharge for 1,552 patients. We represented messages using inferred latent topics, linguistic features (e.g. emotions, activities), and clusters of medical terms. Our analysis indicates that mentioning medication dosage and additional procedures are associated with readmission. Moreover, patients who were readmitted rarely mentioned leisurely activities or described their insights about their health information.

Introduction

Patient portals are secure online websites that healthcare organization provide to grant patients 24 hour access to their health records[1–3]. Portals include a wide range of health information, including discharge summaries, medications, immunizations, and laboratory tests[1,4]. One of the popular functionalities of patient portals is their support of secure messaging between patients and care providers[5,6,7]. Messaging allows patients to keep their healthcare providers informed about their clinical status outside of visits to the clinic and stays in the hospital. Moreover, The use of messaging has been shown to be associated with improved chronic disease management and medication adherence[8,9]. It has been shown that discussions about laboratory test results, reporting new symptoms, and requesting prescription refills are the most common topics in patients' messages[10]. The popularity of these topics demonstrates that patients utilize the portal messages for seeking information from healthcare providers when they are outside of the clinical environment. In addition, patients communicate different needs in their messages, which can be roughly partitioned into logistical (e.g., location of clinic), social (e.g., thanking care providers), informational (e.g., asking about intervention), and medical (e.g., informing physician of a health problem)[11,12]. Notably, it has been shown that more than 70% of messages originating from patients included medical needs[12,13].

There has been limited research into the identification of the symptoms and events that patients communicate through their messages, and their relationship with outcomes such as medication discontinuation or readmission[14,15]. This may be due, in part, to the fact that the information relevant to a patient's clinical status in a message is not explicitly indicated, but rather is in an unstructured form. This makes it essential to first identify potential factors that are indicative of health issues in portal messages, so that healthcare providers can assess and evaluate the health status of a patient. This need is further exacerbated by the rapid growth in the adoption of this technology and the number of patient messages as an artifact[16]. Hence, healthcare organizations would benefit from automating the process of detecting a message that can indicate the risk of a negative outcome, such as readmission.

In this study, we aim to identify the contents and textual features in patient messages that can indicate the possibility of negative outcomes. Specifically, we analyzed the content of messages sent by patients who were diagnosed with ischemic heart disease and hospitalized at Vanderbilt University Medical Center (VUMC). We analyzed the content of messages sent after discharge and within 31 days for two types of patients: 1) those who were readmitted due to an unplanned hospitalization and 2) those who lacked an unplanned hospitalization. We extracted linguistic and textual features, along with patients' demographics, and applied a generalized linear model to learn their association with readmission risk.

Related work

Several studies have analyzed the content of messages sent through patient portals[17,18]. Some of these investigations relied upon manually review, while others applied machine learning and statistical analysis to automatically extract patients' needs and assess the association between the messages and an event of interest[12,13,15]. Certain studies focused on the volume or content of messages and with respect to outcome. For instance, Sulieman et. al. investigated the post-discharge factors that are associated with readmission risk[19]. They found that the number of messages that patients sent after discharge was one of the top predictors of readmission. Yin et. al. extracted the patterns of messaging with healthcare providers, the volume of messages and the content of messages sent by breast cancer patients[15]. The authors combined these features to find associations between messages and the potential for discontinuing hormonal therapy[15]. They observed that mentions of side effects and surgery-related topics were associated with an increased risk of discontinuation. By contrast, they further observed that expressions of gratitude and mentions of drugs prescribed to treat side effects were associated with a decreased risk of discontinuation. North et. al. reviewed and assessed the content of patient messages and its association with the risk of death within 30 days and the risk of hospitalization within 7 days following the message[17]. They found that patients mentioned high risk symptoms in 3.5% of messages and that six hospitalizations (0.09% of messages) were related to a patient message.

Methods

Cohort

We extracted data from the VUMC Synthetic Derivative (SD), a de-identified version of the electronic health record (EHR). We focused our analysis on patients who exhibited ischemic heart disease during an inpatient visit and sent a message through the MyHealthAtVanderbilt (MHAV) patient portal after discharge. We identified the patients who were readmitted within 31 days. We excluded patients who did not send any messages after discharge. We also retrieved patients' age at discharge, gender, race, and ethnicity.

There were 96,044 patients who were diagnosed with ischemic heart disease, with admissions between 1990 and 2018. 6,448 of these patients sent a message using MHAV between 2003 and 2018. The cohort for this study consisted of the 1,552 patients who sent a message within 31 days after the discharge. Of these patients, 40 (2.6%) were readmitted within 31 days. As shown in Table 2, approximately two-thirds of the patients were male and with an average age of 63. The patients were 93% Caucasian and 98% non-Hispanic.

Table 1. Demographics of patients with ischemic heart disease who sent a post-discharge message.

Demographic Feature		**Without readmission N = 1512**	**Readmitted N = 40**
Age		63.0 (12.3%)	62.1 (15.8%)
Gender	Male	1024 (67.7%)	25 (62.5%)
	Female	488 (32.3%	15 (37.5%)
Race	White	1408 (93.1%)	35 (87.5%)
	Black	79 (5.2%)	5 (12.5%)
	Asian	13 (0.9%)	0 (0%)
	Native American	3 (0.2%)	0 (0%)
	Unknown	9 (0.6%)	0 (0%)
Ethnicity	NonHispanic	1487 (98.3%)	39 (97.5%)
	Hispanic	16 (1.1%)	1 (2.5%)
	Unknown	9 (0.6%)	0 (0%)

Message Extraction and Analysis

We extracted the messages exchanged between patients and their healthcare providers, regardless of who initiated the message. For patients who were readmitted, we extracted the messages they sent after discharge and before readmission. For patients without a readmission, we extracted the messages sent within 31 days after the discharge. For patients with multiple admissions, we included the earliest readmission event only.

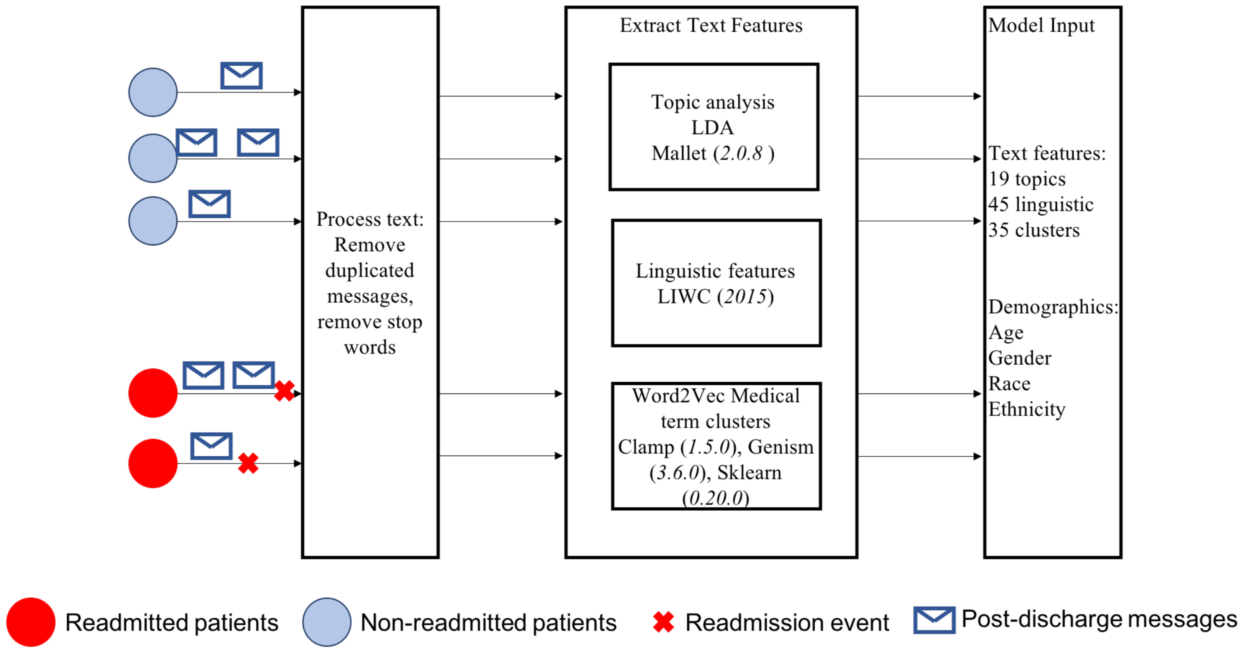


Figure 1. The pipeline for extracting text features from patient portal messages.

Text analysis

We grouped the messages sent by each patient after their discharge into a single document. To profile the content of each patient's message, we preprocessed the messages, performed topic analysis, extracted linguistic features, and defined medical term clusters, as outlined in Figure 1.

Topic Analysis

To extract topics from messages, we applied Latent Dirichlet Allocation (LDA) as implemented in the *Mallet* Java package (version 2.0.8). LDA is a popular topic modeling method that allows a group of documents to be explained by latent topics, each of which can be further explained by the words in the documents. After running LDA, we obtained a topic distribution for each document (e.g., the probability that a document can be explained by each topic), and a word distribution for each topic (e.g., the probability that a topic can be explained by each word). Based on these distributions, together with the lengths of the documents, we calculated the topic distribution across the corpus by combining all the documents into a single large document. LDA has proven effective at summarizing a large amount of text[20]. Since it is an unsupervised machine learning method, we relied on the coherence score to determine the best number of topics. The coherence score is used to measure the extent to which the most probable words in every topic appear together in either the current documents or some external data source (e.g., Wikipedia). A higher coherence score suggests a better topic modeling result. We learned LDA models with 2 to 26 topics (with a step size of 1) and chose the number of topics that exhibit the largest coherence score. To mitigate word sparsity and ensure interpretability, we replaced each term with its lemma form and retained only nouns, verbs, adjectives and adverbs. We also generated the bi-grams of terms using the *genism* python package (version 3.6.0) to capture more meaningful phrases.

Linguistic Features

We applied Linguistic Inquiry and Word Count (LIWC, version 2015) to extract the cognitive, emotional, and social aspects in the messages[21]. LIWC is an effective tool to summarize linguistic features from online generated content[22]. The LIWC package generates approximately 90 linguistic variables, including general descriptor categories (e.g., words per sentence), standard linguistic dimensions (e.g., percentage of pronouns in the message), word categories tapping psychological constructs (e.g., affect, cognition, biological processes, and drives), personal concern categories (e.g., leisure, work, and home), informal language markers (e.g., assents and swear words), and punctuation categories. In our text analysis, we focused on psychological constructs, personal concerns, and informal language. Table 1 shows the 45 features that LIWC extracted.

Table 2. Linguistic features that extracted from the messages.

Linguistic Categories	Features
Affective process	*affect, anger, anxiety, emotion, negative emotion, positive, sad*
Social process	*family, friend, humans, social*
Cognitive process	*cognitive mechanism, cause, certain, discrep, insight, tentative*
Perceptual process	*feel, hear, perception, see*
Biological process	*body, health, ingestion, sexual*
Personal concerns	*work, achieve, leisure, home, money, religion, death*
Relativity	*motion, space, time*
Informal language	*assent, filler, non-fluencies, swear*
Other grammar	*exclusion, inclusion, inhibition, numbers, quantifiers*

Medical Term Clusters

We used *Clamp* (version 1.5.0) to extract the medical terms from the messages, including treatments, clinical problems, and laboratory tests[23]. Given the large number of clinical terms that were extracted, we reduced the dimensionality to represent features more efficiently by grouping terms that shared similar meanings. To do so, we first trained a word2vec model using the clinical communications in the entire SD. We did not use the pretrained word2vec model (e.g., the Google word2vec) because there are many terms (e.g., abbreviations) that do not exist in the pretrained documents. Word2vec generates a vector (or embedding) for each word, where similar words exhibit high semantic similarity based on a cosine function. We trained word2vec using the *genism* python package with a minimum word count of 50, window size of 15, and 100 hidden units.

We retrieved the words' vectors (i.e., word2vec embedding)for each medical term in the messages extracted by Clamp. For the cases where a medical term consists of more than one word, we retrieved the word2vec embeddings for each word and calculated the mean. We applied agglomerative hierarchical clustering with complete linkage, according to a cosine distance, in the *sklearn* python package (version 0.20.0) to cluster the word embeddings. To obtain the most efficient number of clusters, we adopted a metric that generates a number of clusters that is 1) large enough to create efficient and interpretable semantic clusters and 2) small enough to avoid partitioning one cluster into two or more clusters with similar words. This was accomplished through the approach introduced by Yin et. al.[15]. Specifically, we construct clusters where the number of clusters range from 2 to 100 clusters (with a step size of 1). We use the standard deviation of the cluster sizes to determine when to stop the clustering process. Heuristically, the standard deviation tends to become small as the number of clusters increases. To identify the optimal number of clusters, we followed the elbow principle to locate the cluster number where the marginal gain of increasing cluster size begins to diminish.

Content Analysis

We applied logistic regression to evaluate the associations between the content of messages that includes 19 topics, 45 linguistic features, and 35 word2vec clusters and demographics, and the readmission risk. Specifically, we used the Generalized Linear Model (GLM) library in R (version 3.5.2) to learn three association models:

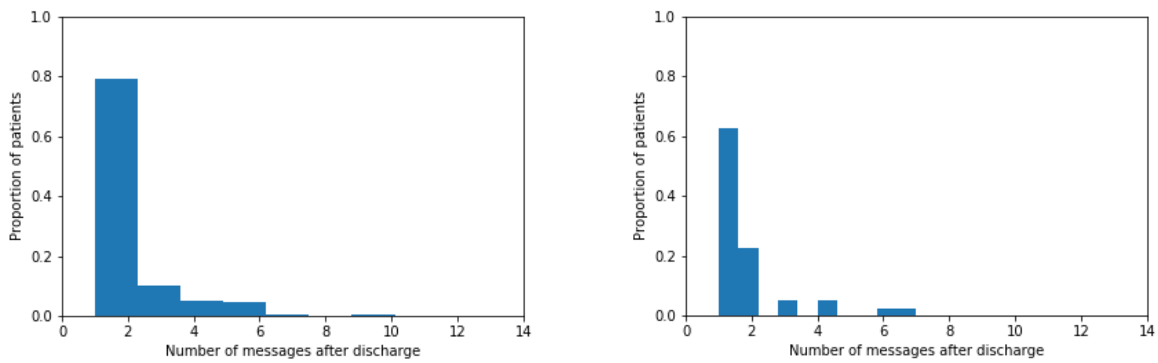
1- Demographics-Only model: This includes age (rescaled to 0-1 range), gender, race, and ethnicity;
2- Message content model: The learned topics (rescaled to [0,1] range), LIWC linguistic features (rescaled to [0,1] range), and medical term clusters; and
3- Demographics and message content model: A combination of models 1 and 2.

For each model, we identified the features with coefficients that were statistically significant at the 0.05 level, their associations, and the Akaike information criterion (AIC) values. The latter is an estimate of the relative quality of statistical models, which is obtained by estimating the amount of information loss in the model. A higher AIC value indicates a lower amount of information loss and thus a better quality.

Results

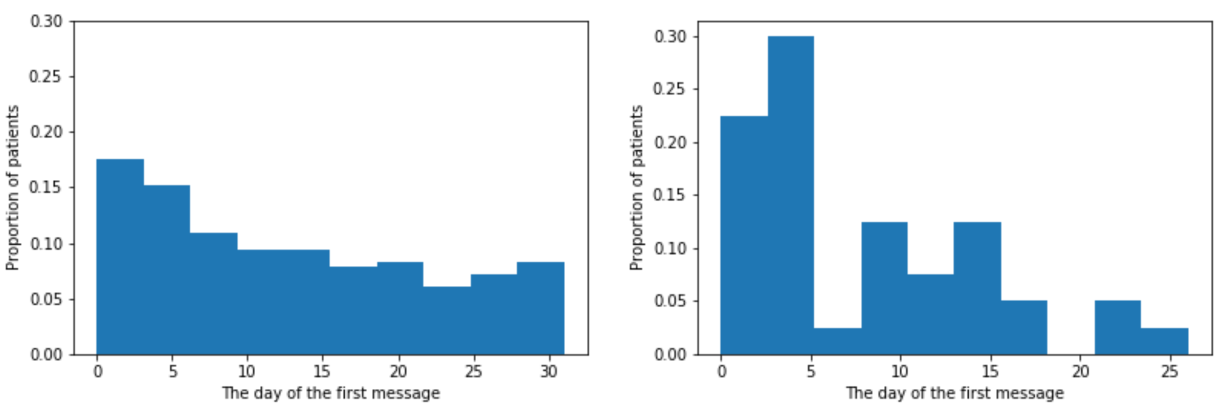
The patients sent 4,052 messages either before readmission or within 31 days after the discharge (for those without a readmission). Figure 2 illustrates the distribution of messages sent by patients according to their readmission status. From the figure, it can be seen that around 60% and 80% of the patients who were readmitted and lacked a readmission, respectively, sent only one message. The average (median) number of messages sent by patients with and without a readmission was 1.75 (1) and 1.87 (1), respectively. We ran Mann-Whitney to evaluate whether the numbers (e.g., number of messages, number of words in messages) were statistically different for patients with readmission and patients without readmission. The difference between the number of messages sent by those two groups was not found to be statistically significant (Mann-Whitney U: statistic=28120, p-value = 0.2). Among the readmitted patients, 20% sent two messages, while only 10% of the non-readmitted patients sent two messages. On average, the messages sent by patients consisted of 700 words overall (median = 381 words), while the readmitted and non-readmitted subgroups consisted of 667 words (median = 546 words) and 700 words (median = 376 words), respectively. The difference between the number of words in messages from patients with and without a readmission was not found to be statistically significant either (Mann-Whitney U: statistic = 27291.0, p-value = 0.15).

More than 50% of readmitted patients sent a message within the first five days after discharge, while 30% of patients who were not readmitted sent a message within 5 days, as depicted in Figures 2(b) and 3(a). Only 10% of patients sent a message five days before the readmission. The mean of the day of the first message sent by a readmitted and non-readmitted patient was 7.8 (SD = 6.3 and median = 5) and 12.9 (SD = 9 and median = 12), respectively. The difference between the day of the first message sent by readmitted patients and non-readmitted patients was statistically significant (Mann-Whitney U: statistic = 20478.0, p-value = 0.0002).



(a) Patients not readmitted within 30 days (b) Patients readmitted with 30 days

Figure 2. Number of messages sent post-discharge sent by patients (a) without and (b) with a readmission.



(a) Patients not readmitted within 30 days (b) Patients readmitted with 30 days

Figure 3. First day a message was sent after discharge by a patient (a) without and (b) with a readmission.

Message Topics

We identified 19 topics in post-discharge messages. Each topic includes a set of words that patients invoked to discuss a particular topic. For instance, when patients ask about, or reschedule, appointments, they use relative phrases such as the *day of the week*, *time*, *schedule*, *confirm*, or *reschedule*. Table 3 lists the most relevant words in each topic ranked by the LDA model. It can be seen that the topics primarily covered appointments, vitals, checking laboratory tests, medications (including prescription and time), and logistics (including discharge locations and communications).

Readmission Risk Associations

The demographic model achieved an AIC of 379; however, none of the features (i.e., age, gender, race, and ethnicity) were statistically significant. In the message content model, we applied 19 topics, 45 linguistic features, and 35 clusters. This model achieved higher quality with an AIC of 462 and contained six features that had a statistically significant association with the readmission risk (as shown in Table 4). We report the statistically significant features. The swear linguistic category was significant and positively correlated with the readmission event. By contrast, both insight and leisure were negatively correlated with the readmission event. Moreover, the words in three clusters listed in Table 5 were significantly associated with readmission. Each cluster includes the words that are similar to each other based on their Word2Vec similarity scores. The words in Clusters 2 and 29 were positively correlated with readmission, while those in Cluster 11 were negatively correlated. It should be noted that removing correlated values from the model did affect the significance of the features. When combining demographic features and content features, we obtained a model with AIC of 468, which was slightly higher than the message content model. In this combined model, only one feature, Cluster 2 was significant. The demographic features were still insignificant in this model.

Table 5 shows the words for each significant word semantic cluster: Cluster 2, Cluster 11, and Cluster 29. We ranked the words in each cluster based on their cosine distance to its centroid (i.e., the mean of the word2vec for the words in the cluster). As Table 5 shows, the most relevant words in Cluster 2 primarily correspond to the medication dosage that was prescribed for the patient. The top words in Cluster 11 are laboratory test names, while Cluster 29 includes the mentions of another clinical event, such as a procedure, laboratory test, and medication.

Table 3. The 30 most relevant words in each of the topics extracted from the messages.

#Topic	Topic	30 Most Relevant Terms	Probability
1	Appointments	*appointment, schedule, appt, dr, follow, week, reschedule, time, hospital, cancel, make, friday, nov, tuesday, clinic, thursday, apr, office, jan, feb, work, aug, advise, discharge, pm, early, confirm, tomorrow, monday, afternoon*	6.53%
2	Monitoring	*weight, day, lasix, morning, today, continue, fluid, send, blood, pressure, potassium, leg, swell, breath, week, daily, increase, feel, foot, advise, dose, lb, shortness, pill, time, afternoon, bumex, mg, yesterday, walk*	5.89%
3	Vital (specifically blood pressure)	*blood, low, pressure, heart, start, send, heart_rate, pulse, feel, record, dr, week, check, normal, back, review, time, monitor, rn, amiodarone, med, rate, make, home, mr, office, episode, today, medication, drop*	5.87%
4	Laboratory test checks	*lab, result, order, draw, blood, week, level, work, clinic, friday, tomorrow, repeat, check, test, dr, day, good, wait, mr, bmp, today, iron, back, home, dose, potassium, low, recheck, send, cbc*	5.80%
5	Time specifically for medication	*mg, daily, tablet, day, medication, increase, med, change, metoprolol, hour, list, hospital, morning, dose, mouth, continue, question, review, make, feb, pressure, follow, dosage, week, tab, add, bid, furosemide, procedure, start*	5.77%
6	Feeling at time	*good, feel, today, hope, morning, great, glad, make, check, talk, tomorrow, yesterday, increase, hospital, start, thing, weekend, week, hear, night, home, feeling, bit, weak, bsn, change, ve, time, wonderful, long*	5.48%
7	Communicatio n	*call, phone, staruser, patient, dr, message, number, back, speak, leave, today, pt, nurse, request, reach, rn, wife, state, return, give, regard, cell, yesterday, notify, result, response, md, provider, miss, set*	5.45%
8	Discomfort feelings and symptoms	*pain, chest, good, leg, time, side, continue, area, leave, incision, problem, arm, normal, walk, due, ms, tylenol, bad, hurt, feel, discomfort, back, drainage, improve, dr, clear, hip, heal, put, level*	5.25%
9	Tests	*blood, test, problem, heart, time, stent, case, cath, cardiologist, symptom, cardiac, month, prior, urine, stress, week, back, result, stress_t, recommend, risk, plan, good, feel, study, put, show, remember, year, plavix*	5.25%

10	Discharge locations	*home, care, mom, health, rehab, daughter, nurse, discharge, jun, hospital, mother, visit, clinic, give, concern, injection, time, today, yesterday, ms, good, question, check, make, institution, vanderbilt, day, discuss, great, stallworth*	5.25%
11	Medication dosage	*inr, warfarin, mg, today, dose, day, coumadin, message, week, start, check, friday, send, cardiac_rehab, institution, oct, back, med, morning, mr, time, clinic, call, pat, result, dr, tonight, increase, work, continue*	5.20%
12	Communication and address	*fax, send, email, mail, phone, form, work, letter, receive, sign, number, office, paperwork, complete, address, copy, dr, melissa, note, fill, time, paper, request, week, feb, write, give, place, mr, day*	5.11%
13	Blood sugar	*pm, unit, blood, sugar, day, check, morning, aug, insulin, bedtime, dose, low, give, time, reaid, high, meal, glucose, dinner, send, lantus, night, week, jul, eat, schedule, start, phone, jun, lunch*	5.08%
14	Treatment	*dr, sleep, night, medication, doctor, day, call, med, give, infection, antibiotic, dialysis, treatment, stop, ray, continue, message, start, problem, treat, care, advise, prescribe, cough, feel, head, aware, discuss, headache, back*	5.03%
15	Appointment time	*work, surgery, return, dr, week, time, month, question, phone, pcp, back, day, issue, follow, request, release, wife, oct, contact, health, place, fax, appointment, part, great, read, care, institution, remove, statin*	4.92%
16	Logistics	*place, information, contact, office, institution, make, order, room, request, medical, report, insurance, note, vanderbilt, find, referral, plan, send, prior, code, infusion, receive, mrs, check, record, hear, additional, cancer, dec, chemo*	4.89%
17	Pharmacy	*prescription, pharmacy, refill, send, vanderbilt, health, detail, link, click, notification, care_provider, script, phone, fax, pick, electronically_sent, supply, plavix, rx, call, fill, prescribe, generic, cvs, street_address, insurance, year, md, kroger, month*	4.71%
18	Communication between clinics and relatives	*send, order, dr, jul, call, pt, oxygen, echo, dad, schedule, make, delegate_relative, put, place, give, message, week, scan, show, today, talk, follow, patient, appt, speak, problem, office, clinic, move, write*	4.66%
19	Medication refill	*day, medication, patient, question, supply_remain, tab, home, contact, phone, answer, call, transplant, address, place, note, device, jun, list, refill, mg, leave, regard, change, prednisone, process, order, make, lpn, mr, message*	3.86%

Table 4. Features with statistically significant beta coefficients in the GLM model based on message content.

Concept	Estimate Std.	Error	z-value	Pr(>|z|)
swear	7.375	3.139	2.350	0.0188
insight	-0.655	0.317	-2.066	0.0389
leisure	-1.893	0.963	-1.966	0.0493
Cluster 2	1.196	0.503	2.377	0.0175
Cluster 11	-2.963	1.257	-2.357	0.0184
Cluster 29	1.404	0.669	2.098	0.0359

Table 5. The top 20 words in each statistically significant cluster. The words are ranked according to their distance from the centroid of the cluster in ascending order.

Cluster 2: Medication dosage	Cluster 11: Laboratory test	Cluster 29: “Another” event
prednisone dosage reduction	labs (cpd cmp ldh	another cxr
current dose diabetes medication	cbc cmp	another mri
higher dose steroids	cbc cmp	another bmp
usual gabapentin dose	cpd cmp ldh uab	another echo
tavr procedure dose	cpd cmp ldh	another dilemn
prednisone dosage	cmp cbc bnp	another treatment
methotrexate dose	cpd cmp igg ig igm spep	another procedure
lortab dosing	cpd cmp	another paracentesis

nebulizer meds	cpd cmp aml	another uti
shot oral dose steroids	cmp cpd	another ultrasound
current prednisone dose	cpd cmp gengraf level	another medication
vancomycin dose	lipids cmp	another ct scan
chemo treatment	lipids cmp	another ct-scan
radio frequency ablation procedure	cpd cmp igg spep serum free light chains code	another cbc
normal asacol dose	cbc cmp ps type screen	another diuretic
current medication regimen	fasting lipids cmp	another ct
paxil dose	cmp ldh	another xray
maintenance dose prednisone	bmp	another infection
prednisone dose	cbc+diff bmp	another antibiotic

Discussion

This investigation yielded several notable findings. First, the message patterns for patients who were readmitted were different from patients who were not readmitted. Specifically, patients who were readmitted tended to send messages earlier than other patients. This might indicate that these patients were experiencing problems or complications after discharge that they communicate to their healthcare providers. Identifying the relevant words, topics, or signals in the message may assist healthcare organizations to identify patients who are at higher readmission risk and, thus, address such complications in a timely manner.

Second, patients' messages included indications regarding patients' health status, health concerns, and social context after discharge, which may be useful for predicting readmission. While conventional features, such as patient demographics did not exhibit a significant association with the readmission status, the model performance was improved by incorporating the message content. We believe this is because it includes patients' activities and concerns. For instance, social information that patients communicated about their leisurely activities, insights, and feelings were statistically significant. The analysis demonstrated that readmitted patients were less likely to write sentences about leisure activities or describe patients' intuitions and insights (e.g., *think* or *know*) as the coefficients of the model indicated. For example, patients who were not readmitted tended to send similar questions, seek feedback or answers, such as "*I think my main concern is how the H/H was trending down at the time of discharge*". One of the messages exchanged between a provider and a patient included a mention of hunting, walking (e.g., "*gone deer hunting", "I walk about half a mile to get my hunting school", "restricted to walking a half hour daily instead of an hour"*). This suggests that the readmitted patients might be experiencing unfamiliar symptoms or feelings, thus limiting their ability to enjoy some social or leisurely activities. Moreover, readmitted patients tended to use swear words in their messages, which may indicate their frustration after discharge.

Third, the medical terms in the model had a significant association with the status of the patient after discharge. Mainly, the messages that have information about the dosage of patient medications and the laboratory tests were significantly associated with readmission. These medical information types exhibited opposing directionality in their association with a patient's outcome. Specifically, the medication dosage had a positive association with readmission, which might indicate that the readmitted patient asked about their medication dosages or explained some side effects with the prescribed dosage. By contrast, the existence of laboratory values in the message had a negative association with readmission. Another significant medical feature is the request to repeat or perform another clinical test or another medication. For example, the mention of an additional clinical event such as *another MRI*, *another CT scan*, *another UTI* have a positive association with readmission. Hence, the necessity of ordering another scan or another test can imply that the patient had complications after the discharge, which increased the probability of readmission.

At the same time, there are several limitations worth noting. First, this a study of a specific population at a single medical center, which calls into question the generalizability of our findings. Second, the dataset size was relatively small. Expanding this analysis to include more phenotypes and a large number of messages could provide more intuition into the associations. Third, we combined the messages sent by a patient, which masks the temporal changes in the topics. Fourth, our text analysis did not handle the negation which we will address in our future work. Fifth, our analysis focused on identifying the indication of readmission in patients messages. We did not evaluate the capability of predicting the readmission using the model. In our future work, we will focus on evaluating the ability of predicting the readmission using the model. Finally, there was a low readmission rate in our dataset. Sending MHAV secure messaging might indicate that the patients used online patient portal to seek advice/information regarding their health, which might explain, in certain degree, the low rate of readmission in MHAV cohort who seeks information in the portal.

Conclusion

Online portals provide a secure channel that allows patients to interact with their healthcare providers. Patients use portal messages to communicate their needs, requests, and questions. While the number of patient portal messages is increasing, analyzing the patient reported information in their messages is still limited. Identifying the signals in portal messages that indicate the risk of readmission can help providers apply interventions to avoid adverse events. This study showed the messages sent by patients with ischemic heart disease after a hospital discharge can be leveraged to predict readmission. The findings specifically showed that leisure activities, intuition (e.g., think), and swear words, as well as medical terms in the messages are associated with readmission events. We believe that future research will benefit by evaluating the capability of our model to predict the readmission using patients message and expanding on the analysis to include other diseases.

Acknowledgements

This research was sponsored in part by NIH grant UL1TR002243.

References

1. HealthIT.gov. What is a patient portal?. [Online]. Available from: https://www.healthit.gov/faq/what-patient-portal [Updated September 29, 2017; Accessed 10 March 2019].
2. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, Stroebel RJ. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health. 2014 Mar 1;20(3):192-8.
3. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. J Med Internet Res. 2014;16(11):e258.
4. Weingart SN, Hamrick HE, Tutkus S, Carbo A, Sands DZ, Tess A, Davis RB, Bates DW, Phillips RS. Medication safety messages for patients via the web portal: the MedCheck intervention. Int J Med Inform. 2008 Mar 1;77(3):161-8.
5. Osborn CY, Rosenbloom ST, Stenner SP, Anders S, Muse S, Johnson KB, Jirjis J, Jackson GP. MyHealthAtVanderbilt: policies and procedures governing patient portal functionality. J Am Med Inform Assoc. 2011 Jul 31;18(Supplement_1):i18-23.
6. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res. 2015;17(2):e44.
7. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc. 2012 Dec 15;20(3):519-25.
8. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (millwood). 2010 Jul 1;29(7):1370-5.
9. Nazi KM, Woods SS. MyHealtheVet PHR: a description of users and patient portal use. AMIA Annu Symp Proc. 2015:1182-1182.
10. Houston TK, Sands DZ, Jenckes MW, Ford DE. Experiences of patients who were early adopters of electronic communication with their physician: satisfaction, benefits, and concerns. Am J Manag Care. 2004 Sep;10(9):601-8.
11. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. Int J Med Inform. 2017 Sep 1;105:110-20.
12. Sulieman L, Gilmore D, French C, Cronin RM, Jackson GP, Russell M, Fabbri D. Classifying patient portal messages using Convolutional Neural Networks. J Biomed Inform. 2017 Oct 1;74:59-70.
13. Cronin RM, Fabbri D, Denny JC, Jackson GP. Automated classification of consumer health information needs in patient portal messages. AMIA Annu Symp Proc. 2015: 1861-70.
14. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk?. J Am Med Inform Assoc. 2013 May 23;20(6):1143-9.

15. Yin Z, Harrell M, Warner JL, Chen Q, Fabbri D, Malin BA. The therapy is making me sick: how online portal communications between breast cancer patients and physicians indicate medication discontinuation. J Am Med Inform Assoc. 2018 Oct 30;25(11):1444-51.
16. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc. 2016 Apr 1;30(4):1432-40.
17. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk?. J Am Med Inform Assoc. 2013 May 23;20(6):1143-9.
18. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, Stewart M, Simon SR. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res. 2014;16(3):e75.
19. Sulieman L, Fabbri D, Wang F, Hu J, Malin BA. Predicting Negative Events: Using Post-discharge Data to Detect High-Risk Patients. AMIA Annu Symp Proc. 2015: 1169-78.
20. Yang SH, Kolcz A, Schlaikjer A, Gupta P. Large-scale high-precision topic modeling on twitter. InProceedings of the 20th ACM SIGKDD international conference on Knowledge Discovery and Data Mining 2014 Aug 24 (pp. 1907-1916). ACM.
21. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. 2015 Sep 15.
22. Yin Z, Chen Y, Fabbri D, Sun J, Malin B.# PrayForDad: Learning the Semantics Behind Why Social Media Users Disclose Health Information. InTenth International AAAI Conference on Web and Social Media 2016 Mar 31. 2016: 456-65
23. Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H, Xu H. CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc. 2017 Nov 24;25(3):331-6.