JAMIA Open, 4(3), 2021, 1–9
doi: 10.1093/jamiaopen/ooab049
Research and Applications

Research and Applications

Classification and analysis of asynchronous communication content between care team members involved in breast cancer treatment

Bryan D. Steitz [1]*, Lina Sulieman[1], Jeremy L. Warner[1,2], Daniel Fabbri[1], J. Thomas Brown[1], Alyssa L. Davis[3], and Kim M. Unertl [1]

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [2]Division of Hematology/Oncology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA and [3]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

*Corresponding Author: Bryan D. Steitz, PhD, Department of Biomedical Informatics, Vanderbilt University Medical Center, 2525 West End Ave., Suite 1475, Nashville, TN 37203, USA; bryan.steitz@vanderbilt.edu

Received 11 January 2021; Revised 6 May 2021; Editorial Decision 12 June 2021; Accepted 16 June 2021

ABSTRACT

Objective: A growing research literature has highlighted the work of managing and triaging clinical messages as a major contributor to professional exhaustion and burnout. The goal of this study was to discover and quantify the distribution of message content sent among care team members treating patients with breast cancer.

Materials and Methods: We analyzed nearly two years of communication data from the electronic health record (EHR) between care team members at Vanderbilt University Medical Center. We applied natural language processing to perform sentence-level annotation into one of five information types: clinical, medical logistics, non-medical logistics, social, and other. We combined sentence-level annotations for each respective message. We evaluated message content by team member role and clinic activity.

Results: Our dataset included 81 857 messages containing 613 877 sentences. Across all roles, 63.4% and 21.8% of messages contained logistical information and clinical information, respectively. Individuals in administrative or clinical staff roles sent 81% of all messages containing logistical information. There were 33.2% of messages sent by physicians containing clinical information—the most of any role.

Discussion and Conclusion: Our results demonstrate that EHR-based asynchronous communication is integral to coordinate care for patients with breast cancer. By understanding the content of messages sent by care team members, we can devise informatics initiatives to improve physicians' clerical burden and reduce unnecessary interruptions.

Key words: workflow, burnout, multidisciplinary communication, breast cancer, electronic health records

INTRODUCTION

Managing care for patients with cancer requires communication and coordination among numerous specialists and team members who are often distributed by clinic location.[1–6] Electronic health record (EHR)-based asynchronous clinical messaging has emerged as a primary technology to support team-based communication.[7–10] EHR-based messaging is characterized by a centralized structure, which supports messages sent by an individual to a team of providers and staff.[10–13] In this format, multiple individuals in pre-determined

LAY SUMMARY

Clinical messaging is important for communication among healthcare providers and staff to ensure that patients receive high-quality and up-to-date care. However, managing the high volume of messages requires extensive work, which can distract providers from clinical duties. In order to develop solutions to reduce physicians' work of managing messages, it is necessary to first understand the types of information that are communicated through these messages. To identify these information types, we applied natural language processing to classify the content of nearly 2 years of messages, sent by individuals treating patients with breast cancer, from the electronic health record. We found that the majority of messages (63.4%) contained logistical information, while only 21.8% contained clinical information. Administrative staff sent the majority of messages containing logistical information and physicians sent the majority of messages containing clinical information. Our results highlight that these messages are important for the planning and organization of healthcare and that administrative staff are essential to the process of coordinating care.

teams, often by clinic affiliation, can review and respond to each message.[10]

In theory, a team-based approach to clinical messaging allows individuals to respond to their respective messages while ensuring that the rest of the team remains informed about the patient's care.[7,14] However, in practice, many of these messages lead to unnecessary interruptions from an individual's current scope of patient care.[15] A growing research literature has suggested that care team members receive an increasingly high volume of asynchronous clinical communications, which lead to professional exhaustion and burnout.[11,16–19] However, recent studies have highlighted the wider task of managing the inbox and message triage as a particular source of work.[20,21] A study by Arndt et al[20] of family physicians found that managing the messaging inbox takes 23% of their workday. In our previous work, we evaluated the scope and volume of clinical messages sent between care team members, including the actions performed on those messages.[5,6] We found that over half of the messaging actions performed by a care team treating patients with breast cancer involved only reading a message without subsequently responding to the message. Effective message triage supports the opportunity to improve message response time and reduce care team work.[22]

Asynchronous clinical communication has become integral to support effective care delivery, but a meaningful process to triage the large volume of clinical communications is necessary to improve care team workload. To date, studies to assess clinical message content have primarily been conducted using patient portal messaging.[13,23,24] These studies have applied a variety of natural language processing (NLP) approaches, including regression, decision trees, random forests, and neural networks to identify the needs that patients' communicate in their messages. A study by Sulieman et al[24] found that in a sample of 3000 patient-generated messages, 642 contained only logistical or social information that may not require a physician to respond. Our previous work has found that patients with breast cancer themselves were involved in only 26.8% of message threads in their EHRs, suggesting that a large volume of messages cannot be classified using models previously developed with a patient communication focus.[5] In this work, we apply NLP to the clinical communications among clinical team members providing care to patients with breast cancer at an academic medical center. We aim to discover, quantify, and describe the distribution of message content, including analysis of care team member roles and message time.

METHODS

We conducted this study at the Vanderbilt-Ingram Cancer Center at Vanderbilt University Medical Center (VUMC). VUMC is an academic medical center located in middle Tennessee and provides referral care across the southeastern United States. VUMC includes a 758-bed Vanderbilt University Hospital and receives 1.6 million annual ambulatory visits.[25] During the period of this study, providers and staff at VUMC used an institutionally developed EHR, StarPanel, for all clinical functions—including secure messaging.[10] The Vanderbilt University Institutional Review Board approved this study (Protocol 160843).

Study population and data sources

Our study population included any patient who had an appointment with a VUMC-affiliated medical or surgical breast oncologist between January 1, 2015, and November 1, 2017, and was the subject of at least 1 message thread.[4,26] We extracted all EHR-based secure asynchronous message logs corresponding to patients in our cohort between November 1, 2016, and November 1, 2017. Message log data included a unique employee identifier, a unique patient identifier, a message thread identifier, and the timestamp of the message. We mapped each employee identifier to their job role and grouped job roles into 5 classifications: administrative staff, clinical staff, oncology providers, noncancer-specific providers, and other.[4,5] Clinical staff included clinical technicians, nurses, nurse practitioners, and physician assistants who were involved in direct patient care. Staff classified as "other" included individuals such as pharmacists, pharmacy technicians, volunteers who were neither involved in clinical administrative tasks nor direct patient care. We identified provider specialty using their national provider identifier. We defined medical oncologists, surgical oncologists, plastic surgeons, and radiation oncologists as oncology providers due to the frequency with which they are involved in the treatment of patients with breast cancer.[5]

Taxonomy

We developed a sentence-level classification scheme of care team communication as shown in Box 1. We adapted our taxonomy from the parent categories of the Taxonomy of Consumer Health Information Needs[23] with modifications to reflect communication types between providers and staff as identified by informal interviews with VUMC Breast Center providers and staff. The taxonomy contains 4 primary categories that identify the informational purpose of each message: Clinical Information, Medical Logistics Information, Nonmedical Logistics Information, and Social Information. Content that did not fit into any of the 4 primary categories were classified as "Other." While each message could contain multiple communication types, individual sentences were required to be labeled with a single type.

Box 1. Taxonomy of care team communication types

A. Clinical Information: Information involving clinical reasoning or delivery of medical care
B. Medical Logistics Information: Information involving the coordination or scheduling of medical care
C. Nonmedical Logistics Information: Communications about pragmatic information that is not related to medical care (eg, location of a clinic or a copy of a medical record)
D. Social Information: Communications related to social interactions or an interpersonal relationship that is not directly related to any of the above needs
E. Other: Communication that does not fit into one of the above categories

Gold standard

To create a gold standard training set, we randomly selected 200 message threads from our dataset. Messages were split into sentences and subsequently deidentified using the MITRE Identification Scrubber Toolkit.[27] Each message was independently annotated using the Taxonomy of Care Team Communication Types by two annotators familiar with clinical medicine. Annotators reviewed the messages through an electronic crowdsourcing interface where they were organized by thread and divided into sentences.[28] Annotators labeled each sentence with a single communication type. We measured interrater reliability with Cohen's kappa, and it was 0.38. Following independent coding, two additional annotators (BDS and KMU) manually reviewed annotations and resolved discrepancies through discussion until consensus was achieved. For each annotation that did not match from the original independent annotators, we kept the consensus annotation from discussion between the two additional annotators.

NLP approach

We built and evaluated five multiclass machine learning classifiers to identify communication types in secure messages between providers and staff. Machine learning classifiers included: (1) random forest; (2) multinomial naïve Bayes; (3) support vector machine (SVM); (4) bidirectional encoder representations for transformers (BERT);[29] (5) clinical BERT, a BERT model that was previously trained on Medical Information Mart for Intensive Care clinical notes;[30] and (6) SciBERT, a BERT model that was trained on a corpus of scientific literature.[31] We used the cased models for BERT, clinical BERT, and SciBERT—each of which were accessed through the HuggingFace Transformers library.[32] For both BERT models, classification was performed using a linear classification layer that sits atop the BERT architecture. Each classifier output a categorical classification corresponding to 1 of the 5 classifications in our taxonomy for each sentence. We identified the optimal model parameters for random forest, naïve Bayes, and SVM classifiers using grid search in sci-kit learn during the training phase for each respective model.[33] Similarly, we tuned the BERT, Clinical BERT, and SciBERT models using the training set of our communications dataset. In tuning the BERT models, we added a fully connected layer atop the BERT architecture while freezing all other layers. We additionally trained a linear classification layer to categorize sentences into one of the five information categories. Optimal parameters for each model were subsequently applied the test dataset for final model evaluation.

Features that served as inputs to our random forest, naïve Bayes, and SVM classifiers included bag of words (BoW), term frequency-inverse document frequency (TF-IDF), and a Word2Vec model pretrained on a Google News dataset. We preprocessed the messages by removing nonalphanumeric characters and excluding stop words retrieved from the Natural Language Toolkit Python package.[34] We represented the corpus of messages as a matrix in which each *sentence* in a message corresponds to a row and features are represented in designated columns. We used the words' counts in each sentence for BoW representation. For TF-IDF representation, each sentence was represented by a TF-IDF value that ranged between 0 and 1 and was calculated on the training set. We averaged each word's vector to obtain Word2Vec representations. All classifiers were trained and tested on the gold standard corpus of 200 message threads that included 2074 sentences with 5-fold cross-validation.

Statistical analysis

We evaluated the performance of each classifier and respective feature selection method using one-versus-all area under the receiver operator curves (AUCs), micro and macro F1 scores, accuracy, precision, and recall. We chose the model with the highest F1 score to classify the sentences in our entire dataset (ie, the unannotated messages). We compared the distribution for the predicted labels in the entire dataset to the gold standard labels. We similarly combined sentences for each respective message to determine concept co-occurrence, which we visualized using an UpSet graph.[35] We combined annotations per message by taking the union of sentences' annotations for the respective message.

We calculated descriptive statistics to evaluate message concepts relative to care team member role and activity. First, we analyzed the content and messages sent and received by care team member role. Second, we summarized the volume of each message content classification sent between roles. Finally, we compared message content by oncology provider clinic activity and by working hours. We determined clinic activity by days in which a provider had scheduled appointments or procedures. Working hours were defined as any time spent on the EHR-based secure messages between 7:00 am and 7:00 pm local time.[18,20]

RESULTS

Our gold standard set contained 200 unique message threads consisting of 2074 sentences in 766 messages—a median of 3 sentences per message. The sentence-level annotations contained 568 (27%) medical logistics, 486 (23%) social, 411 (20%) nonmedical logistics, 346 (17%) clinical information, and 263 (13%) other information. Using the gold standard, we developed, trained, and optimized 5 classification algorithms (Table 1). BERT-base yielded the highest accuracy (tied with Clinical BERT), macro F1 score, micro F1 score, and AUC with the values 0.72, 0.72, 0.7, and 0.91, respectively.

Hence, we subsequently applied BERT-base to the full dataset of clinical communications sent about patients in our cohort between November 1, 2016, and November 1, 2017. The full dataset contained 613 877 sentences across 81 857 unique messages. These messages were sent by 4044 unique care team members about 3766 patients (Table 2). Across all roles, more messages contained logistical information (63.4%) than any other classification. Similarly, 30.2% of all messages sent by cancer providers included at least 1 sentence containing clinical information. We present an UpSet visualization in Figure 1 of sentence-level classification sets and their respective co-occurrence in clinical messages.

Table 3 presents the content of messages sent between care team member roles. Administrative staff sent more messages to other administrative staff (44.4%) than care team members of any other role. Similarly, clinical staff and physicians sent the most messages to other clinical staff. Clinical staff and physicians sent more medical logistics information than any other information classification, regardless of recipient role. There were 20 174 messages sent by care team members that contained only social information or information classified as “Other,” of which 16 985 ended a message thread. A total of 5784 of these messages were sent by cancer providers, representing 52.7% of the total threads in which cancer providers were involved.

There were 21 providers in our network who we classified as directly related to breast cancer treatments. These providers sent 15 912 messages through 10 970 distinct threads. Table 4 presents oncology provider messaging statistics by time of day and clinic activity. Each cancer provider sent an average of 13.6 messages (standard deviation [SD] = 11.6) on days with scheduled clinic activity compared to 9.9 messages (SD = 5.9) when they did not have scheduled clinical duties. Regardless of time and clinical activity, medical logistics information was the most common type of sent information, occurring in 52.4%–55.3% of all sent messages. On days in which providers did not have clinical activity, 69.8% of messages received after hours contained clinical information.

Table 1. Classification model metrics

Classifier	Optimal parameters	Parameter range	Accuracy	Macro-precision	Macro-recall	Micro-F1	Macro-F1	AUC
Random forest	Maximum depth = 100	1–150 in increments of 2	0.59 (0.047)	0.62 (0.053)	0.54 (0.053)	0.61 (0.047)	0.72 (0.051)	0.74 (0.029)
(SD)	Maximum features = 2	1–100 in increments of 1						
	Number of estimators = 50	1–200 in increments of 2						
	Feature selection method = Word2Vec	BoW, TF-IDF, Word2Vec						
Naïve Bayes	Alpha = 0.5	0.1–1.5 in increments of 0.1	0.59 (0.026)	0.68 (0.049)	0.61 (0.026)	0.65 (0.026)	0.63 (0.032)	0.78 (0.016)
(SD)	Feature selection method = BoW	BoW, TF-IDF, Word2Vec						
Support vector machine	Penalty = 0.1	0.1–5.0 in increments of 0.1	0.61 (0.036)	0.66 (0.044)	0.64 (0.039)	0.68 (0.036)	0.65 (0.041)	0.8 (0.023)
(SD)	Regularization = L2	L1, L2						
	Tolerance for stopping criteria = 1.3	0.1–2.0 in increments of 0.1						
	Feature selection method = Word2Vec	BoW, TF-IDF, Word2Vec						
BERT base	Epochs = 2	1–10 in increments of 1	0.72 (0.023)	0.7 (0.022)	0.7 (0.019)	0.72 (0.023)	0.7 (0.023)	0.91 (0.017)
(SD)	Learning rate = 3e−5	1e−5, 2e−5, 3e−5, 4e−5, 5e−5						
	Max sequence length = 128	8–256 in increments of 8						
Clinical BERT	Epochs = 2	1–10 in increments of 1	0.72 (0.026)	0.77 (0.030)	0.65 (0.055)	0.69 (0.023)	0.64 (0.026)	0.89 (0.023)
(SD)	Learning rate = 3e−5	1e−5, 2e−5, 3e−5, 4e−5, 5e−5						
	Max sequence length = 128	8–256 in increments of 8						
SciBERT	Epochs = 3	1–10 in increments of 1	0.71 (0.030)	0.7 (0.031)	0.69 (0.032)	0.71 (0.032)	0.69 (0.022)	0.90 (0.016)
	Learning rate = 3e−5	1e−5, 2e−5, 3e−5, 4e−5, 5e−5						
	Max sequence length = 128	8–256 in increments of 8						

AUC: area under the receiver operator curve; BERT: bidirectional encoder representations for transformer; BoW: bag of words; SD: standard deviation; TF-IDF; term frequency-inverse document frequency.

DISCUSSION

In this study, we assessed and described the content of secure asynchronous messages exchanged between providers to coordinate treatment for patients with breast cancer. We trained and applied NLP classification algorithms to discover message content sent by all care team members treating a cohort of patients over one year. There have been other studies to investigate clinical message content, but these studies have primarily focused on messages originating from patients through the patient portal.[23,24,36–39] These studies have applied both manual[37,39] and automated classification techniques.[23,24,36,38] A study by North et al[37] used manual review to identify that 3.5% of patient portal messages contained urgent, high-risk, clinical needs. Another study by Cronin et al compared NLP approaches to apply the taxonomy of consumer health information needs[23] to patient portal messages.[36] They found that 72.3% and 24.8% of studied patient portal messages contained medical information and logistical information, respectively. However, in our previous work, we found that patients are involved in only 26.8% of message threads. To the best of our knowledge, this is one of the first studies to automatically classify the content of secure EHR-based clinical messages sent between care team members, across all care team roles.

Our analysis was supported by NLP-based classification methods, which we trained using a gold standard set of messages. We

Table 2. Care team messaging statistics by care team member role

	Administrative staff	Clinical staff	Physician (cancer provider)	Physician (noncancer specialist)	Other	Total
Number of care team members	1214	1661	21	972	176	4044
Number of patients	3623	3675	3766	2354	2236	3766
Number of message threads	25 664	34 532	10 970	11 761	2246	51 157
Number of sent messages	48 087	65 619	15 912	16 458	2906	148 982
Clinical information (%)	5941 (12.4)	15 076 (23.0)	4802 (30.2)	5956 (36.2)	710 (24.4)	32 485 (21.8)
Medical logistics (%)	28 619 (59.5)	35 340 (53.9)	8540 (53.7)	7697 (46.8)	1597 (55.0)	81 793 (54.9)
Nonmedical logistics (%)	20 790 (43.2)	25 743 (39.2)	3724 (23.4)	3963 (24.1)	1170 (40.3)	55 390 (37.2)
Social information (%)	13 945 (29.0)	18 613 (28.4)	7815 (49.1)	5926 (36.1)	1139 (39.2)	47 438 (31.8)
Other (%)	8221 (17.1)	16 608 (25.3)	4545 (28.6)	4448 (27.0)	439 (15.1)	34 261 (23.0)
Number of received messages	32 968	50 175	11 404	12 158	1735	10 8441
Clinical information (%)	3792 (11.5)	12 504 (24.9)	4314 (37.8)	4707 (38.7)	409 (23.6)	25 726 (23.7)
Medical logistics (%)	21 155 (64.2)	27 376 (54.6)	7003 (61.4)	6701 (55.1)	966 (55.7)	63 201 (58.3)
Nonmedical logistics (%)	11 855 (36.0)	21 458 (42.8)	3633 (31.9)	4294 (35.3)	534 (30.8)	41 774 (38.5)
Social information (%)	12 160 (36.9)	15 163 (30.2)	4906 (43.0)	3738 (30.7)	691 (39.8)	36 658 (33.8)
Other (%)	6413 (19.5)	10 633 (21.2)	2558 (22.4)	2843 (23.4)	430 (24.8)	22 877 (21.1)

* Since messages can contain multiple sentences, percentages for sent and received message content will sum to greater than 100%.

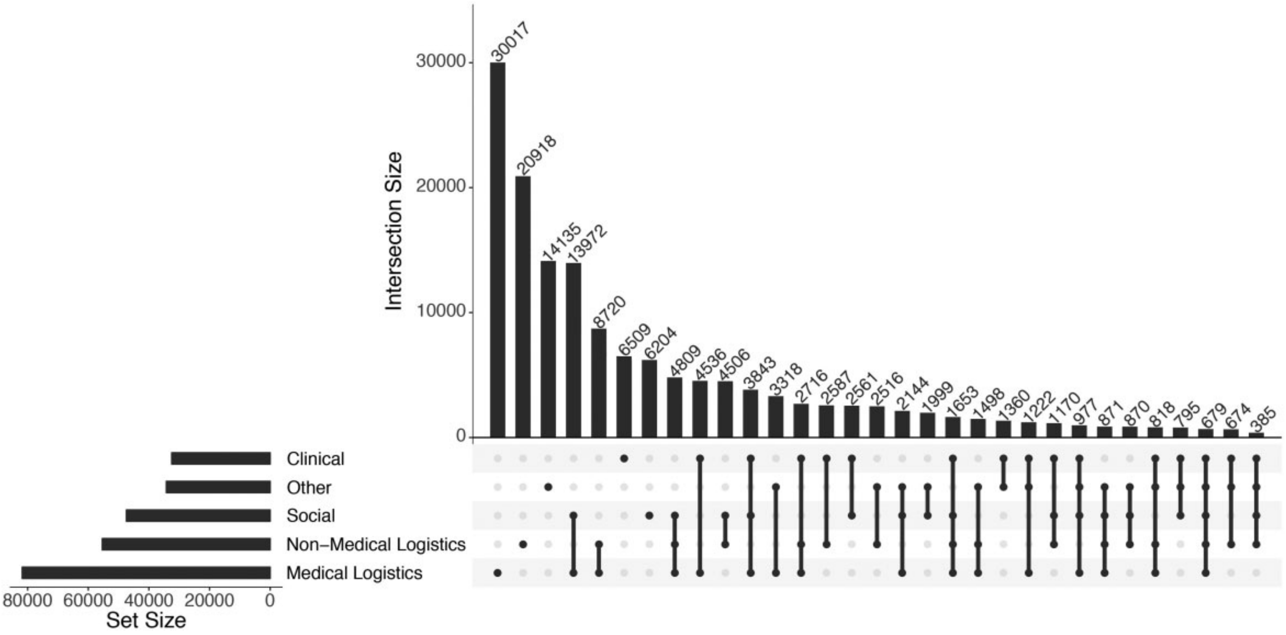

Figure 1. UpSet Visualization of Messages Grouped by Classification. The bar graph in the lower left corner depicts sentence-level distribution across each category. Each row in the dot graph represents a classification category; solid dots represent each category part of the intersecting sets. The center bar graph depicts the number of messages in each intersection.

compared multiple classification models and feature types. The best classifier had high predictive ability and was able to determine which categories of information were present in a sentence. We found that messaging was a primary work product of breast cancer care coordination, such that care team members performed messaging actions in 37.5% of all EHR sessions, averaging 29.8 messaging sessions per day.[40] Automated classification of asynchronous messages may aid in informatics initiatives to reduce messaging load, such as through message triage or by identifying nonurgent messages that do not require immediate notification.

Our NLP approach, however, was subject to several limitations. First, our classifiers were trained on a limited gold standard set of 200 message threads containing 766 unique messages. Previous work suggests that our classification performance may increase with a larger gold standard corpus.[23,36] However, we were able to improve our classification performance using BERT for transfer learning, which reflects findings from previous studies.[41] The gold standard corpus from which we trained and tuned the NLP models had a relatively low Cohen's kappa score. However, during a manual review by the independent adjudicators, we noted that many of the discrepancies were related to slight differences in text selection approaches. We hypothesize that annotations differed, in part, due to the differing degrees of clinical experience between reviewers. Nonetheless, we conducted a second manual review with 2 experi-

Table 3. Content of messages exchanged between care team roles

	Administrative staff	Clinical staff	Physician (cancer provider)	Physician (noncancer specialist)	Other
Administrative staff					
Clinical information (%)	1214 (8.5)	2357 (20.1)	427 (16.0)	451 (23.1)	233 (14.6)
Medical logistics (%)	7913 (55.1)	6023 (51.4)	1189 (44.4)	814 (41.7)	700 (43.8)
Nonmedical logistics (%)	5359 (37.3)	3858 (32.9)	406 (15.2)	436 (22.3)	376 (23.5)
Social information (%)	4077 (28.4)	2893 (24.7)	1330 (49.7)	705 (36.1)	1023 (64.1)
Other (%)	2707 (18.9)	3505 (29.9)	914 (34.1)	525 (26.9)	439 (27.5)
Total number of messages	14 359	11 727	2677	1952	1597
Clinical staff					
Clinical information (%)	755 (7.9)	3209 (18.2)	1409 (29.3)	1837 (32.3)	1262 (29.2)
Medical logistics (%)	5758 (60.4)	8382 (47.6)	2223 (46.2)	2437 (42.9)	1824 (42.2)
Nonmedical logistics (%)	2946 (30.9)	8014 (45.5)	1011 (21.0)	1146 (20.2)	1150 (26.6)
Social information (%)	3015 (31.6)	4261 (24.2)	2147 (44.6)	1699 (29.9)	2744 (63.4)
Other (%)	1798 (18.9)	4295 (24.4)	1367 (28.4)	1642 (28.9)	1212 (28.0)
Total number of messages	9535	17 625	4809	5685	4327
Physician (cancer provider)					
Clinical information (%)	323 (9.5)	997 (21.0)	562 (28.0)	142 (36.1)	248 (33.7)
Medical logistics (%)	2115 (62.4)	2544 (53.7)	910 (45.3)	173 (44.0)	331 (45.0)
Nonmedical logistics (%)	871 (25.7)	1876 (39.6)	556 (27.7)	77 (19.6)	173 (23.5)
Social information (%)	1134 (33.5)	1566 (33.0)	897 (44.6)	198 (50.4)	452 (61.5)
Other (%)	691 (20.4)	1326 (28.0)	672 (33.4)	97 (24.7)	192 (26.1)
Total number of messages	3390	4741	2009	393	735
Physician (noncancer provider)					
Clinical information (%)	203 (7.9)	1410 (25.1)	183 (39.6)	573 (30.6)	489 (44.8)
Medical logistics (%)	1416 (55.0)	2778 (49.5)	185 (40.0)	722 (38.6)	448 (41.0)
Nonmedical logistics (%)	1001 (38.9)	2441 (43.5)	80 (17.3)	445 (23.8)	240 (22.0)
Social information (%)	567 (22.0)	1341 (23.9)	251 (54.3)	523 (28.0)	766 (70.1)
Other (%)	497 (19.3)	1489 (26.5)	160 (34.6)	820 (43.8)	347 (31.8)
Total number of messages	2576	5614	462	1871	1092
Other					
Clinical information (%)	399 (13.3)	2962 (29.0)	600 (42.1)	1076 (48.1)	209 (30.8)
Medical logistics (%)	1700 (56.5)	4973 (48.6)	640 (44.9)	964 (43.1)	311 (45.9)
Nonmedical logistics (%)	832 (27.6)	3168 (31.0)	283 (19.9)	591 (26.4)	210 (31.0)
Social information (%)	1029 (34.2)	3661 (35.8)	784 (55.1)	878 (39.3)	368 (54.3)
Other (%)	784 (26.0)	3251 (31.8)	400 (28.1)	663 (29.7)	185 (27.3)
Total number of messages	3011	10 227	1424	2236	678

Row-wise care team member roles represent the role from which a message was sent. Each column represents the role of provider who received the respective message. The heatmap visualizes the percent of each information type.

enced researchers to discuss discrepancies and determine consensus annotations. Interestingly, we saw decreased performance from the original BERT model when we applied pretrained BERT models trained on clinical notes and scientific text.[30,31] We noted a similar decrease in performance during our preliminary work comparing a Word2Vec model trained on Google news and a Word2Vec trained on PubMed articles and clinical notes.[42] We hypothesize that clinical notes and scientific text contain a larger degree of clinical detail and jargon, which is reflected in our results suggesting that 47% of the sentences contained logistical information, compared to only 17% that contained clinical information. Unlike the ClinicalBERT and BERT-base models which utilize the original vocabulary built on nondomain-specific text, SciBERT incorporated a scientific domain-specific vocabulary, which we hypothesize could also affect performance due to the lack of clinical information contained within our message corpus. Our features did not account for grammar or other sentence-level semantics; it is unclear whether performance would be improved with the addition of these higher-level features. Additionally, we train, tested, and applied our NLP algorithms at the sentence-level of each message. As a result, it is likely that many sentences in the same message were split between the training and test datasets, making it possible to memorize features about the overall message resulting in an overestimate of model performance. However, we hypothesize that there is minimal semantic dependence between sentences, which we will test in future work. Additionally, our cross-validation included only training and testing datasets without an additional validation dataset. We made this decision to maximize the amount of data available for model development. We note the lack of a separate validation set to measure model generalizability as a limitation to our approach. Future work will aim to develop a larger corpus gold standard messages on which to apply our classification algorithms. Using a larger gold standard corpus will allow us to explore more granular information types such that we can further understand message content with the goal of improving message triage tasks.

We focused our analysis on patients who had at least one appointment with a breast medical or surgical oncologist at our institution. We chose this patient population such that we could understand the full scope of message content sent by a care team treating patients with breast cancer over a one-year period. How-

Table 4. Oncology provider messaging statistics by time and clinic activity

	In clinic			Not in clinic		
	Working hours	After hours	Total	Working hours	After hours	Total
Number of sent messages	11 136 (93.5%)	778 (6.5%)	11 916	3633 (90.9%)	363 (9.1%)	3996
Clinical information (%)	3289 (29.5)	251 (32.3)	3540	1149 (31.6)	113 (31.1)	1262
Medical logistics (%)	6006 (53.9)	430 (55.3)	6436	1905 (52.4)	199 (54.8)	2104
Nonmedical logistics (%)	2726 (24.5)	199 (25.6)	2925	729 (20.1)	70 (19.3)	799
Social information (%)	5364 (48.2)	379 (48.7)	5743	1871 (51.5)	201 (55.4)	2072
Other (%)	3216 (28.9)	250 (32.1)	3466	985 (27.1)	94 (25.9)	1079
Number of received messages	7891 (94.4%)	471 (5.6%)	8362	2790 (91.7%)	252 (8.3%)	3042
Clinical information (%)	1167 (14.8)	97 (20.6)	3050	1088 (39.0)	176 (69.8)	1264
Medical logistics (%)	4791 (60.7)	294 (62.4)	5085	1771 (63.5)	147 (58.3)	1918
Nonmedical logistics (%)	2482 (31.5)	165 (35.0)	2647	896 (32.1)	90 (35.7)	986
Social information (%)	3398 (43.1)	195 (41.4)	3593	1193 (42.8)	120 (47.6)	1313
Other (%)	1728 (21.9)	102 (21.7)	1830	671 (24.1)	57 (22.6)	728

* Since messages can contain multiple sentences, percentages for sent and received message content will sum to greater than 100%.

ever, previous studies have found that patients with breast cancer receive care from multiple healthcare institutions. Inter-institution collaborations are not often supported by EHR-based messaging and require other means of communication. Many care coordination activities occur through synchronous communication (eg, phone calls, in-person conversations).[3,43,44] As a result, our findings cannot capture all communication among care team members treating patients with breast cancer, or across all organizations involved in their care. Similarly, we also do not account for other forms of synchronous and asynchronous means to support provider communication within our institution (eg, email). However, during our study period at VUMC, EHR-based asynchronous communication was the preferred means of communication as a way to document conversation among care team members.[10]

Understanding the information discussed in clinical messages is a critical first step to recognizing opportunities to reduce messaging workload. Across all team member roles, we found that 63.4% of messages discussed logistical information. Similarly, all roles sent and received more medical logistics information than any other information type. These results suggest that EHR-based asynchronous clinical communication is highly important in coordinating care, although it is not clear if it is the most efficient or effective approach to care coordination or if the best-qualified people are being asked to deal with these messages. We also found that the 81% and 80% of all logistical information were sent and received by administrative and clinical staff, respectively. This indicates the importance of staff in these roles to coordinate care, which reflects results from our previous work and the importance of including all care team members when evaluating care coordination analyses.[5] Numerous previous studies have related clerical and administrative work, such as responding to messages, as a major factor in physician burnout.[17,45] We hypothesize that systematically classifying messages to identify messages that can be answered by other care team members can help to triage messages and reduce physicians' messaging workload. We also found that physicians send and receive more messages containing clinical information than team members of any other role. Nonetheless, these communications accounted for less than 40% of all messages. We hypothesize that providers utilize other forms of communication to communicate more urgent needs.

Our results indicate that 11% of sentences were classified as "Other." In our manual review of messages, we found that the majority of these sentences contained an acknowledgment of a previous message. Similarly, we found that there were 16 985 messages that contained only sentences classified as social or "Other" information that ended a message thread. There were 5784 of these messages that were sent by cancer providers, representing 52.7% of the total threads in which these providers were involved. These results suggest that there is an opportunity to support functionality that can predict the end of a message thread and marking the thread as resolved. Future work could seek to develop algorithms to automatically detect these completed threads without requiring unnecessary messaging actions and responses.

Numerous studies have suggested that work outside of normal working hours and on days without clinic responsibility leads to professional burnout.[17,45,46] In our analysis of cancer provider messaging by clinic activity and time, we found that there continued to be a large amount of messaging activity performed outside of direct clinical responsibility. We found that despite clinical activity and time of day, logistical information persists as the most common type of information. However, our results indicate that nearly 70% of received messages after hours when cancer providers did not have scheduled clinical activity contained clinical information. Nonetheless, only 31% of the sent messages contained clinical information. We hypothesize that cancer providers triage these messages based on urgency. Future work should seek to develop algorithms to predict message urgency, which could reduce unnecessary notifications for nonurgent messages.

CONCLUSIONS

Our study demonstrates that EHR-based asynchronous communications are integral to coordinating the care of patients with breast cancer. This study is one of the first to apply NLP to classify the content of messages sent between care team members. Understanding the content of messages sent by care team members affords the opportunity to devise informatics initiatives to improve physicians' clerical burden and reduce unnecessary interruptions.

FUNDING

BDS and JTB were supported by the 4T15LM007450 training grant from the United States National Library of Medicine.

AUTHOR CONTRIBUTIONS

BDS, LS, and KMU were involved in the conception of the work, data analysis and interpretation, drafting of the article, and revision of the article. JLW and DF provided guidance in data analysis and interpretation and participated in drafting the article. JTB and ALD were involved in data collection and participated in drafting the article.

ACKNOWLEDGMENTS

The authors would like to thank Joseph Coco for his assistance in enrolling study participants and managing the annotation platform.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY STATEMENT

The data underlying this article cannot be shared due to the identifiable content contained within clinical messages.

REFERENCES

1. Saini KS, Taylor C, Ramirez AJ, *et al.* Role of the multidisciplinary team in breast cancer management: results from a large international survey involving 39 countries. *Ann Oncol* 2012; 23 (4): 853–9.
2. Haynes K, Ugalde A, Whiffen R, *et al.* Health professionals involved in cancer care coordination: nature of the role and scope of practice. *Collegian* 2018; 25 (4): 395–400.
3. Easley J, Miedema B, Carroll JC, *et al.* Coordination of cancer care between family physicians and cancer specialists: importance of communication. *Can Fam Physician* 2016; 62 (10): e608–15.
4. Steitz BD, Levy MA. A social network analysis of cancer provider collaboration. *AMIA Annu Symp Proc* 2016; 2016: 1987–96.
5. Steitz BD, Unertl KM, Levy MA. Characterizing communication patterns among members of the clinical care team to deliver breast cancer treatment. *J Am Med Inform Assoc* 2020; 27 (2): 236–8.
6. Steitz BD, Levy MA. Evaluating the scope of clinical electronic messaging to coordinate care in a breast cancer cohort. *Stud Health Technol Inform* 2019; 264: 808–12.
7. Coiera E. Clinical communication: a new informatics paradigm. In: *Proceedings of the AMIA Annual Fall Symposium*; 1996; 17. American Medical Informatics Association.
8. Parker J, Coiera E. Improving clinical communication: a view from psychology. *J Am Med Inform Assoc* 2000; 7 (5): 453–61.
9. Katz SJ, Moyer CA. The emerging role of online communication between patients and their providers. *J Gen Intern Med* 2004; 19 (9): 978–83.
10. Giuse DA. Supporting communication in an integrated patient record system. *AMIA Annu Symp Proc* 2003; 2003: 1065.
11. Tai-Seale M, Dillon EC, Yang Y, *et al.* Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. *Health Aff (Millwood)* 2019; 38 (7): 1073–8.
12. Unertl KM, Weinger MB, Johnson KB, *et al.* Describing and modeling workflow and information flow in chronic disease care. *J Am Med Inform Assoc* 2009; 16 (6): 826–36.
13. Cronin RM, Davis SE, Shenson JA, *et al.* Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. *Appl Clin Inform* 2015; 06 (02): 288–304.
14. Agarwal R, Sands DZ, Schneider JD. Quantifying the economic impact of communication inefficiencies in U.S. hospitals. *J Healthc Manag* 2010; 55 (4): 265–81; discussion 281–2.
15. Gregory ME, Russo E, Singh H. Electronic health record alert-related workload as a predictor of burnout in primary care providers. *Appl Clin Inform* 2017; 8 (3): 686–97.
16. Lieu TA, Altschuler A, Weiner JZ, *et al.* Primary care physicians' experiences with and strategies for managing electronic messages. *JAMA Netw Open* 2019; 2 (12): e1918287.
17. Adler-Milstein J, Zhao W, Willard-Grace R, *et al.* Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *J Am Med Inform Assoc* 2020; 27 (4): 531–8.
18. Overhage JM, McCallie D. Physician time spent using the electronic health record during outpatient encounters. *Ann Intern Med* 2020; 172 (3): 169–7.
19. Hilliard RW, Haskell J, Gardner RL. Are specific elements of electronic health record use associated with clinician burnout more than others? *J Am Med Assoc* 2020; 12: 573–10.
20. Arndt BG, Beasley JW, Watkinson MD, *et al.* Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med* 2017; 15 (5): 419–26.
21. Kane B, Sands DZ; for the AMIA Internet Working Group, Task Force on Guidelines for the Use of Clinic-Patient Electronic Mail. Guidelines for the clinical use of electronic mail with patients. *J Am Med Inform Assoc* 1998; 5 (1): 104–11.
22. Haun J, Hathaway W, Chavez M, *et al.* Clinical practice informs secure messaging benefits and best practices. *Appl Clin Inform* 2017; 8 (4): 1003–11.
23. Cronin RM, Fabbri D, Denny JC, *et al.* Automated classification of consumer health information needs in patient portal messages. *AMIA Annu Symp Proc* 2015; 2015: 1861–70.
24. Sulieman L, Gilmore D, French C, *et al.* Classifying patient portal messages using convolutional neural networks. *J Biomed Inform* 2017; 74: 59–70.
25. Vanderbilt University Medical Center Factsheet. Nashville, TN: Vanderbilt University Medical Center; 2018. https://prd-medweb-cdn.s3.amazonaws.com/documents/patientandvisitorinfo/files/Factsheet_2018_v29_web.pdf
26. Danciu I, Cowan JD, Basford M, *et al.* Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform* 2014; 52: 28–35.
27. Aberdeen J, Bayer S, Yeniterzi R, *et al.* The MITRE Identification Scrubber Toolkit: design, training, and assessment. *Int J Med Inform* 2010; 79 (12): 849–59.
28. Ye C, Coco J, Epishova A, *et al.* A crowdsourcing framework for medical data sets. *AMIA Jt Summits Transl Sci Proc* 2018; 2017: 273–80.
29. Devlin J, Chang M-W, Lee K, *et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019: 4171–86.
30. Alsentzer E, Murphy J, Boag W, *et al. Publicly Available Clinical BERT Embeddings*. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019: 72–8.
31. Beltagy I, Lo K, Cohan A. SciBERT: a pretrained language model for scientific text. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 2019: 3615–20; Hong Kong, China.
32. Wolf T, Debut L, Sanh V, *et al.* Transformers: state-of-the-art natural language processing. *ArXiv Comput Lang* 2019; 1–8.
33. Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; 12: 2825–30.
34. Loper E, Bird S. *NLTK*. Morristown, NJ, USA: Association for Computational Linguistics; 2002: 63–70.
35. Lex A, Gehlenborg N, Strobelt H, *et al.* UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014; 20 (12): 1983–92.

36. Cronin RM, Fabbri D, Denny JC, *et al.* A comparison of rule-based and machine learning approaches for classifying patient portal messages. *Int J Med Inform* 2017; 105: 110–20.
37. North F, Crane SJ, Stroebel RJ, *et al.* Patient-generated secure messages and eVisits on a patient portal: are patients at risk? *J Am Med Inform Assoc* 2013; 20 (6): 1143–9.
38. Sulieman L, Robinson JR, Jackson GP. Automating the classification of complexity of medical decision-making in patient-provider messaging in a patient portal. *J Surg Res* 2020; 255: 224–32.
39. Robinson JR, Valentine A, Carney C, *et al.* Complexity of medical decision-making in care provided by surgeons through patient portals. *J Surg Res* 2017; 214: 93–101.
40. Steitz BD, Unertl KM, Levy MA. Quantifying electronic health record usage time among breast cancer care teams to manage asynchronous clinical messages. *Appl Clin Inform*.
41. Peng Y, Yan S, Lu Z. *Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets*. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019: 58–65.
42. Zhang Y, Chen Q, Yang Z, *et al.* BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Sci Data* 2019; 6 (1): 1–9.
43. Graetz I, Reed M, Rundall T, *et al.* Care coordination and electronic health records: connecting clinicians. *AMIA Annu Symp Proc* 2009; 2009: 208–12.
44. Gorin SS, Haggstrom D, Han PKJ, *et al.* Cancer care coordination: a systematic review and meta-analysis of over 30 years of empirical studies. *Ann Behav Med* 2017; 51: 1–15.
45. Gardner RL, Cooper E, Haskell J, *et al.* Physician stress and burnout: the impact of health information technology. *J Am Med Assoc* 2019; 26 (2): 106–14.
46. Saag HS, Shah K, Jones SA, *et al.* Pajama time: working after work in the electronic health record. *J Gen Intern Med* 2019; 34 (9): 1695–2.