



Emergency care and the patient experience: Using sentiment analysis and topic modeling to understand the impact of the COVID-19 pandemic

Sharon Chekijian¹ · Huan Li² · Samah Fodeh³

Received: 7 June 2021 / Accepted: 2 August 2021 / Published online: 13 August 2021
© IUPESM and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The COVID-19 pandemic has presented many unique challenges to patient care especially in emergency medicine. These challenges result in an altered patient experience. Patient experience refers to the cumulative impression made on patients during their medical visit and is measured by a standardized survey tool. Patient experience is considered a key measure of quality of care. The volume of survey data received makes it difficult to spot trends and concerns in patient comments. Topic modeling and sentiment analysis are well documented analytic techniques that can be used to gain insight into patient experience and make sense of vast quantities of data. This study examined three periods of time, pre, during and post-COVID-19 first wave in order to identify key trends in sentiment and topics related to patient experience. Previously collected, anonymized Press Ganey (PG) survey data was used from three northeastern emergency department that make up an academic emergency department. Data was collected for three contiguous time periods: Pre-COVID-19 (12/10/2019- 3/10/2020), During COVID-19: (3/11/2020–6/10/2020), and Post-first wave COVID-19 (6/11/2020- 9/10/2020). Preprocessing of the data was carried out then a sentiment label (i.e., positive, negative, neutral, mixed) was assigned by the tool. These labels were used to assess the validity of Press Ganey labels. Next, a topic modeling approach from machine learning was used to analyze the contents of the patient comments and uncover concerns and perceptions of patient experiences. Themes that emerged from the analysis of patient comments included concerns over personal safety and exposure to the virus, exclusion of family from decision making and care and high levels of scrutiny over systems issues, care, and treatment protocols. Topic modeling showed shifting priorities and concerns throughout the three periods examined. Prior to the pandemic, patient comments were largely positive and focused on technical expertise and perceptions of competence. New topics and concerns that patients reported relevant to the pandemic were identified during-COVID-19. Comments on systems issues regarding processes to limit viral spread and concerns over family/visitor restrictions were dominant. Although there was evidence of praise and appreciation of the efforts of staff there was also a high level of scrutiny of the processes encountered during the emergency visit. Sentiment analysis and topic modeling offer a unique method for organizing and analyzing the shifting concerns of patients and families. Suggestions of interventions are made to address these evolving concerns. The automation of analysis using artificial intelligence would allow for rapid and accurate analysis of patient feedback.

Keywords Patient experience · Machine learning · COVID-19 · Emergency medicine · Sentiment analysis · Thematic analysis

1 Introduction

Patient experience refers to the cumulative impression made on patients during their medical visit. It is measured by a standardized survey tool and is considered a key measure of quality of care. [1] The strongest correlates to improved patient experience in emergency medicine are decreased wait times and boarding, being informed about delays, and a sense by the patient that staff cared about them as a

This article is part of the COVID-19 Health Technology: Design, Regulation, Management, Assessment

✉ Sharon Chekijian
Sharon.chekijian@yale.edu

Extended author information available on the last page of the article

person. [2] Communication is a key component of patient experience be it interpersonal or informational to relay visit expectations and results. [2, 3] The strongest correlates to improved patient experience in emergency medicine are decreased wait times and boarding, being informed about delays, and a sense by the patient that staff cared about them as a person. [4] Addressing patient experience in emergency medicine is a daunting task with the inherent stresses of an unplanned visit, an acute illness, and doctors and staff unfamiliar to the patient. Patient and family control over the visit is at a minimum and experience scores and comments often reflect this discomfort. [2] The COVID-19 pandemic has further strained this already challenging paradigm. Survey data including comments present an opportunity to understand patient experience and perspectives related to their visits and gain insight into their shifting concerns. This study investigates the level of patient satisfaction, as well as their concerns and perceptions by analyzing the narrative feedback collected as part of the Press Ganey (PG) survey tool. The survey tool was distributed per protocol to patients who were treated and released from the emergency departments at two large urban and one community site of a large academic medical center. This study utilizes machine learning approaches to analyze PG narratives and identify patients concerns and perspectives pre, during and post the first wave of the COVID-19 pandemic. Sentiment analysis and topic modeling were applied to uncover themes and

topics that patients communicate in the narrative portion of the survey. Key insight is gained into the shifting concerns of emergency patients in regard to the pandemic. Within the system, the first patient with confirmed SARS-CoV-2 infection was treated on March 10, 2020. Comments for three, three-month periods: pre, during and post first wave of the COVID-19 pandemic are examined. The three time periods examined are framed by the index case.

2 Background

In prior studies, free-text comments from patient experience surveys have been qualitatively and quantitatively analyzed using both manual and machine learning methods such as sentiment analysis and topic modeling. Several qualitative studies have attempted to characterize the contents in patient comments and responses to better understand their concerns and improve quality of health services. Lopez et al. [5] downloaded patient reviews from online rating websites and identified global themes including: overall excellence, negative sentiment, and professionalism. Specific factors were identified, for example interpersonal manner, technical competence, and system issues. (Table 1) Other elements that fall under these categories were also reported. Analyzing the sentiment of the patients comments most reviews were rated positive (63%).

Table 1 Taxonomy of patients' satisfaction themes from Lopez et al. [5]

Five Global Themes				
Overall Excellence	Recommendation	Negative Sentiment	Professionalism	Intent not to Return
Twenty-Nine Specific Factors				
Interpersonal Manner		Technical Competence		System Issues
• Empathic		• Knowledgeable		• Appointment Access
• Friendly		• Detailed		• Appointment Wait Time
• Helpful		• Efficient		• Practice Environment
• Trustworthy		• Clinical Skills		• Practice Health Information Technology
• Time Spent During Appointment		• Follow-up		• Practice Location
• Put at Ease		• Referrals		• Cost of Care
• Listens		• Perceived Poor Decision Making		• Negative View of Healthcare
• Explains		• Perceived Success of Treatment		• Method of Physician Selection
• Longevity of Relationship with Clinician		• Complementary / Alternative Medicine		

A sentiment analysis of six physician rating websites by Emmert et al. found that around 70% of ratings were positive. [6] A more rigorous evaluation of the sentiment of patient comments was conducted by Ellimootil et al. [7] who manually reviewed urology rating sites and classified the sentiments as extremely positive, positive, neutral, negative and extremely negative. Most ratings (75%) were extremely positive, positive, or neutral. Doyle, et al. [8] extracted topic categories by identifying search terms for a meta-analysis of patient experience. Terms were divided into relational and functional aspects. “Relational aspects” include interpersonal manner, emotional and psychological support, patient-centered decisions, clear information, and transparency. “Functional aspects”, matched Lopez et al.’s themes of professionalism, technical competence, and systems issues. These themes include effective treatment, expertise, clean environment, and coordination of care.

Machine learning approaches have been used to perform quantitative analysis via topic modeling and sentiment analysis of patient comments. Maramba, et al. [9] analyzed word frequencies in post-appointment surveys. Surveys were divided into favorable and unfavorable ratings. They found words like “surgery”, “excellent”, “service”, “good”, and “helpful” were the five most distinctive words from satisfied patients. The words “doctor”, “feel”, “appointment”, “rude”, and “symptoms” were the five most distinctive words used by dissatisfied patients.

Doing-Harris et al. [10] developed a classifier to detect topics mentioned in patient survey responses. They evaluated their classifier using a schema of 28 topics that they developed and manually annotated in 300 responses. The topic modeling approach revealed complaints about appointment access, appointment wait, and time spent with physician [10]. Brody et al. applied a topic modeling approach as well based on Latent Dirichlet Allocation to 33,654 online reviews of different types of practitioners and identified words associated with both specialty-independent themes (e.g., recommendation, manner, anecdotal) and specialty specific themes (e.g., general practitioner: prescription and tests, dentist: costs, obstetrician/gynecologist: pregnancy). [11] Topic modeling was also utilized to indicate areas for quality improvement. [12, 13] Sentiment analysis was applied to patient responses to automatically determine the positive or negative polarity of the comment. Greaves et al. obtained 81%, 84%, and 89% agreement between quantitative ratings of care and those derived from free-text comments using sentiment analysis for cleanliness, being treated with dignity, and overall recommendation of hospital respectively. [14]

This study builds on existing research and seeks to characterize and understand overall patients concerns via comments submitted pre, during and post first wave COVID-19

using both sentiment analysis and thematic and topic modeling. The analysis reveals priorities pre, during the pandemic and the new realities and concerns encountered after the first wave of the pandemic.

3 Materials and methods

3.1 Press ganey survey data

Previously collected, anonymized PG survey data was used from three emergency department locations that comprise the network of a northeastern academic emergency department. The index case of SARS-CoV-2 infection was treated on March 10, 2020. Data was collected for three contiguous time periods as follows: pre-COVID-19 (12/10/2019-3/10/2020), during-COVID-19: (3/11/2020–6/10/2020), and post-first wave COVID-19 (6/11/2020- 9/10/2020). The three sites are staffed by the same group of physicians, residents, advanced practice providers and nurses. Survey comments from all three locations were aggregated. The PG survey includes questions about patient experience in various categories including demographics, arrival, nurses, care provider, tests, family or friend, personal insurance information, personal issues, and overall assessment. A sentiment label (i.e., positive, negative, neutral, mixed) was assigned by the tool to the free-text in each one of these categories based on sentiment analysis. Comments with the “mixed” label were excluded from the dataset for two reasons: They represent a small portion of the entire dataset and the study focused on comments likely to contribute to further understanding.

3.2 Validation of sentiment labels of patients comments

The PG analytical methods used to label free-text comments as being positive, negative or neutral is proprietary. PG classifies comments into three categories: people, places and process, each of which have separate topic areas, such as physicians and nurses, etc. [15, 16] These topics can be further broken down into sub-topics, such as “listen” and “knowledgeable.” These sub-topics are then labeled as being positive or negative and the overall comments are then scored according to their positivity or negativity. The lack of information regarding the exact methods used by PG tool to detect the polarity of a comment (positive, negative or neutral) led the team to manually re-label the patient comments. Patient comments were double-coded by two of the authors and labeled as positive, negative or neutral. Specifically, a sample of 50 comments were labeled by two of the authors. Inter-annotator agreement was computed and was 90%. A third author resolved discrepancies and the annotators met to discuss and refine their coding strategy. Inter-annotator

agreement was rechecked every 50 comments and double-coded to ensure that agreement didn't degrade as is the standard practice for text studies [17–19]. With this procedure interrater agreement increased to 96%.

3.3 Data preprocessing

The comments for each period were preprocessed separately as per standard protocols. Using the natural language processing tool (NLTK) [20], each comment was tokenized into its constituting words. Punctuation was removed and words were converted into lower case letters. Stop words, i.e. common words in English that don't necessarily contribute to understanding the embedded topics in the free-text, were removed from the final list of words/tokens. Most frequent and rare words are considered "noise" and are removed before modeling. However, deciding a threshold at which frequent/rare words are dropped depends on the nature of the dataset. A Pareto chart was used to plot the words and help determine the appropriate cut-off point. A Pareto chart is a graph that indicates the frequency of words, as well as their cumulative impact. For example, if the focus is on the words that constitute 80% of the total frequency, words with frequency < n (say 3) will be dropped from the analysis. After applying the threshold to the words, bigrams were composed from the remaining words. Each period pre, during, and post-COVID-19 had its unique list of unigrams and bigrams.

3.4 Sentiment analysis validation

The sentiment labels of the free-text comments in the PG survey as often used as the gold standard to develop classification models that can automatically detect the polarity of the comments. [6, 10] In order to validate the Press Ganey labels used as a basis for the thematic and topic modeling, the process is described below. Due to the lack of information about the approach used to generate these labels, our study undertook a validation process using the gold standard labels generated by the manual annotations assigned by the coders/authors. The sentiment of patient comments was investigated in the three periods pre, during and post-COVID19 to assess the impact of the pandemic on patient experience during the emergency visit. Using the gold standard sentiment labels that were manually assigned to patient comments by the study team, positive, negative and neutral comments were enumerated for all periods individually and trends were compared over time. The PG labels were validated using our gold standard labels to align the labels with our comprehension of patient comments. Sensitivity, precision and F1-score were computed to evaluate PG performance. F1-score is the harmonic mean of precision and sensitivity.

3.5 Topic modeling

Topics describe the subjects of a text. Topic modeling estimates latent topics. A topic modeling approach from machine learning was used to analyze the contents of the patient comments and uncover concerns and perceptions of patient experiences. Topic modeling, namely Latent Dirichlet Allocation (LDA) [21] was employed to extract the hidden topics and concerns reported in the free text responses. LDA [21] is a probabilistic modeling algorithm based on the generative paradigm of text that uses observed variables (words) to estimate hidden variables (topics). Unigrams (terms with single words) were used to estimate topics as well as n-grams (terms with multiple words). N-gram analysis is usually used to add context to the topics. The main idea of LDA is that documents are represented as random mixtures over latent topics where each topic is a distribution over terms. It produces two matrices: the topic versus term matrix and the comment versus topic matrix. The number of latent topics is unknown in prior and needs to be estimated empirically. Through manual analysis of a range of alternative models with different number of topics, the number of topics was identified that yielded the most semantically coherent and distinct topics, compared with specifications with more or fewer topics. Topics were manually investigated to see whether they reflected the themes encountered during manual annotation. A coherence score was computed for each topic model to assess the quality of the learned topics. The coherence score is computed for each topic to measure the degree of semantic similarity between words in the topic. This score helps distinguish between topics that are semantically interpretable and topics that are artifacts of statistical inference. The same analysis was carried out for the three time periods: pre, during and post-COVID-19.

The gensim implementation of LDA topic modeling was used. [22, 23] The default values for all hyper-parameters including Alpha and Beta were used. Comments were parsed and tokenized into words for each period using the gensim package and Natural Language Toolkit (NLTK). The coherence scores were calculated to validate the number of the topics chosen. The MALLET topic model package default function was used.

Topic validation is key to assessing whether the substantive meanings of the topics and their words are parallel with the qualitative meanings in the patients' comments. Past work has advocated the use of sample documents to validate the substantive meaning of each topic. In this study, the top ten comments associated with each topic were examined to validate the topic's substantive meaning. Determination of the top comments per topic was based on ranking topics within the comment versus topic matrix produced by the model. Themes were assigned to the identified topics

Table 2 Description of the patient comments dataset

Time Period	Number of Comments	Number of Comments (without mixed labels)	Number of Words (Unigrams + Bigrams)
Pre-COVID19 (12/10/2019–3/10/2020)	1,719	1,658	797
During COVID19 (3/11/2020–6/10/2020)	1,670	1,490	950
Post-COVID19 (6/11/2020–9/10/2020)	3,017	2,742	1,027
Total	6,406	5,890	2,774

following the Lopez et al.[5] categorization taxonomy shown in Table 1. A new theme “visit activity” was created to map a topic that was consistently surfacing while experimenting with different topic models. It contained words describing activities in the room or while being admitted. It includes words like room, hallway, brought, waiting-area, bed, back, call, blood-pressure, front-desk, water, information. In addition, a new general theme was defined called “COVID-19” under which specific factors that emerged due to the pandemic were listed.

The major concerns and perceptions of patients and how they changed throughout the pandemic were identified using topic modeling. Weights were assigned to the identified topics in each time period and ranked for comparative purposes. Weights were computed using the comment versus topic matrix which encodes the proportion of each topic within a comment. The matrix was dichotomized by applying a threshold (t) to each cell in the matrix. Within a comment, topics with proportions $< t$ are set to 0 and those with proportions $> t$ are set to 1. T was set equal to the average of all cell values in the matrix. The weight of a topic was then computed as the number of comments that have that topic (proportion = 1) divided by the total number of comments in a given period. The weight of a topic quantifies the prevalence of that topic in that period. Topics were ranked based on their weights. This analysis was repeated for all three time periods.

4 Results

The original dataset had a total of 6,406 comments. 516 comments that had “mixed” labels were removed leaving 5,890 comments that were analyzed. Comments were

parsed and tokenized into words for each period. Words that accounted for 80% of the total frequency using the Pareto graph were retained and bigrams were computed. A total of 2,774 words were retained after applying the 80% cut-off including 122 bigrams. (Table 2).

5 Pre, During, and Post-COVID-19 Sentiment Analysis

Based on manual annotations there was a total of 3,791 positive comments, 1,718 negative comments and 897 neutrals. The distribution of these comments across the three periods pre, during, and post-COVID-19 is shown in Table 3. Pre-COVID-19 80% of patients’ comments were labeled positive compared to 53% and 51% in the periods during and post-COVID-19. The percent of negative sentiment was 12, 33% and 32% respectively. (Table 3).

While the absolute number of comments with positive sentiments decreased during the first wave of COVID-19 (During COVID-19 period), it increased in June entering the post-COVID-19 first wave. The number of comments overall increased. The absolute number of comments with negative sentiment continued to increase as COVID-19 started in March 2020 through October 2020. (Fig. 1).

5.1 Reporting validation of PG sentiment labels using the study gold standard

The overall performance of PG sentiment label is impressive and alleviates concerns about using PG labels as basis for the study of sentiment analysis. This was true throughout all three time periods examined. Sensitivity, precision and F1-score were 0.86, 0.94 and 0.89, respectively. The positive

Table 3 Coded sentiment labels based on “gold standard”

Time Period	Positive		Negative		Neutral		Total
	Count	(row %)	Count	(row %)	Count	(row %)	
Pre-COVID19 (12/10/2019–3/10/2020)	1,377	(80.10)	203	(11.81)	139	(8.09)	1,719
During COVID19 (3/11/2020–6/10/2020)	881	(52.75)	554	(33.17)	235	(14.07)	1,670
Post-COVID19 (6/11/2020–9/10/2020)	1,533	(50.81)	961	(31.85)	523	(17.33)	3,017
Total	3,791		1,718		897		6,406

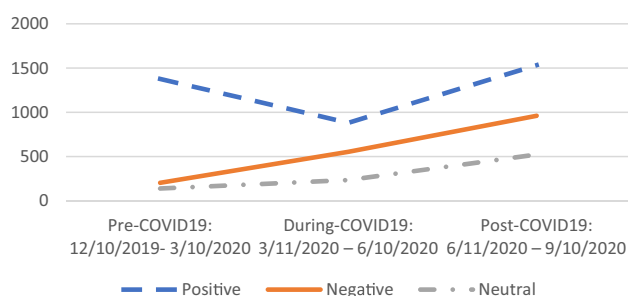


Fig. 1 Trends of sentiments across the three periods: pre, during, and post COVID-19

class was detected with 0.99, 0.96, 0.97 sensitivity, precision and F1-score compared to 0.96, 0.94, 0.95 of the negative class. Out of all positive comments, 0.05% (14 + 3/3,650) were deemed negative or neutral based on PG (Table 4). An acceptable margin of error of 4% (26 + 28/1532) in detecting the negative comments was observed; 26 and 28 comments were mistakenly labeled as positive and neutral, respectively. The least reliable performance of the PG tool was observed with the neutral comments as 39% (200 + 77/708) were mistakenly labeled as positive (200 comments) and negative (77 comments). The corresponding sensitivity, precision and F1-score are 0.61, 0.93, and 0.74, respectively. (Table 4).

5.2 Pre, during and post-COVID-19 thematic and topic analysis

Discreet models were generated with different total numbers of topics (5,10,15,25) and coherence scores for each model were computed in order to assess quality of the learned topics. There was a difference in scores between 5 and 25 topics. For example, an increase from 0.36 to 0.55 was observed in pre-COVID-19, 0.44 to 0.48 during COVID-19 and finally 0.4 to 0.41 in the post-COVID-19 period. Increasing the number of topics beyond 25 resulted in a small increase, if any, in the cohesive score but also generated topics that were either repetitive or with no specific theme. Therefore, models were selected with 25 topics. These 25 topic models revealed many of the themes in Lopez schema that were also encountered during manual annotation, as well as new topics that emerged due to COVID-19. The list of all topics is presented

Table 4 Validation of the Press Ganey sentiment labels. Inter-rater agreement rate = 96%

Gold Standard (manually annotated)	Positive	Negative	Neutral	Total
Positive	3,633	14	3	3650
Negative	26	1478	28	1532
Neutral	200	77	431	708
Total	3,859	1569	462	5800

in Table 5. The new theme “visit activity” showed up in all three periods. The general themes described in Lopez taxonomy including overall excellence, negative sentiment, professionalism, and recommendation were captured via LDA topic modeling analysis in the pre-COVID-19 period.

The same themes were detected in the during and post-COVID-19 periods with the exception that the negative sentiment theme was included under the COVID-19 theme, as shown in .

Table 5. We included nine topics under the COVID-19 theme. Six of those topics were labeled using the Lopez taxonomy risk factors but included COVID-19 related words like: “Corona, COVID, mask, sanitize, virus, pandemic, exposed,” and “due to COVID”. These topics include interpersonal skills (listens), technical competence (knowledgeable), negative sentiment (no family allowed), technical competence (knowledgeable), overall excellence, system issues/practice environment. The newly emergent topics include: communication with family containing words like (“family, phone, husband, wait-car, due to COVID, Corona, security”), delayed wait for results (“COVID-19, long, waiting, delayed, results, CT-scan, test, informed, 2 h”) satisfaction with staff despite stress (“pandemic, staff, stressful, anxiety, performed, impressed, exposed, wearing”) and finally protocols concerns (“COVID, patient, mask, wearing-mask, face, shield, properly, waiting-room, uncomfortable”).

For a closer look at the words constituting these topics word clouds are included for each topic in Fig. 2. The words with the largest font are the most important in a given topic. All these sub-topics appeared in patient comments in the during COVID-19 period except the protocols concerns sub-topic which showed up in the post-COVID-19 period which aligns with national polices encouraging the public to follow these protocols and mitigation strategies.

Table 6 shows the prevalent topics in each period ranked based on their weights. The top nine topics for each period are presented. Pre-COVID-19 patient comment themes are mostly dominated by satisfaction with the technical competence of the staff clinical skills and the treatment they received. Intent to recommend the hospital was also evident in their responses. Negative sentiment was also present but ranked as the 8th topic.

6 Discussion

The practice environment worldwide changed drastically amidst the pandemic. Operational and public health measures were actualized during the COVID-19 first wave and continued in the post-COVID-19 period such as mask wearing, social distancing, screening, temperature taking. In addition, in-hospital measures such as visitor restriction, and limitation of testing and treatment even admission based

Table 5 Topics and themes identified in the periods: pre-, during, and post-COVID-19. The number of identified high level topics is 9. Each topic is represented by a different color. Subtopics are displayed for each topic

Topic/Theme	Pre-COVID-19	During COVID-19	Post-COVID-19
appointment activity	X	X	X
interpersonal manner/ empathy	X		X
interpersonal manners/ explains	X	X	X
interpersonal manner/ friendly	X		X
interpersonal manners/ helpful	X	X	X
interpersonal manners/ time spent	X		
interpersonal manners /listens			X
interpersonal manners/ put at ease			X
negative sentiment	X		
overall excellence	X	X	X
professionalism	X	X	X
recommendation	X	X	
system issues/ appointment wait time	X	X	X
system issues/ practice environment (cleanness)	X		
system issues/ practice environment (privacy)	X	X	
system issues/ staff	X	X	
system issues/ appointment access	X	X	
system issues / practice health technology			X
system issues/ practice environment			X
technical competence/ clinical skills	X	X	
technical competence/ efficient	X		X
technical competence/ knowledgeable	X	X	
technical competence/ successful treatment	X	X	X
technical competence /efficient	X		X
technical competence / detailed			X
COVID-19/ communication with family		X	
COVID-19/ interpersonal manners (listens)		X	
COVID-19/ long wait for results		X	
COVID-19/ negative sentiment (no family allowed)		X	X
COVID-19/ overall excellence		X	X
COVID-19/ satisfaction with staff despite stress		X	
COVID-19/ technical competence (knowledgeable)		X	
COVID-19 / protocols concerns			X
COVID-19/ system issues/ practice environment			X

on institutional criteria and guidelines were unfamiliar to patients. Pressure on staff to minimize interaction times and exposure to known or suspected positive patients made serial communication and treatment difficult and, in some cases, awkward. Patient interviews were sometimes done by phone or alternative device from outside the patient room. Given the pressure on the healthcare system, protocols were developed to determine if patients were able to be discharged. Patients were sometimes discharged even when they felt unwell because oxygen levels did not meet criteria for admission. Non COVID-19 patients were similarly discharged home at unprecedented rates whenever appropriate discharge follow-up was available. Difficult decisions such as these take more discussion and precautionary discharge

instructions, intensifying the need and desire for communication at a time where it is perhaps perilous to do so.

These new practice environment realities were transmitted in the comments analyzed in this study. New topics emerged in the period during COVID-19. The communication between healthcare providers and the patient's family appeared in many comments. For instance, many patients complained or commented that their family members had to wait in the parking lots to await updates.

During COVID-19, dissatisfaction with staff and administrative procedures during the visit as well as dissatisfaction with wait-time were ranked in the top 5 themes. While some patients praised the staff for their professionalism in handling stress caused by COVID-19, others criticized their

Pre-COVID19



During-COVID19



Post-COVID19



Fig. 2 Words clouds representing topics characterizing pre, during, and post-COVID-19 periods

treatment during intake drawing attention to faulty processes. Public health measures reinforced in the community seem to have raised the bar for care within the emergency department. Expectations surrounding hygiene and social distancing were heightened and when they were not met, the comments reflected this discrepancy. Patients seemed to pay more attention than before to the facility's COVID-19 practice environment and procedures, such as perceived lack of hygiene in the bathrooms or patient care areas, proximity to other patients, hallway boarding and other processes that were discordant with ideal pandemic management. (Table 6, column 2, topics 1,7) The reported reduction of the waiting time due to decreased patient volumes during COVID-19 was captured in routinely tracked departmental statistics but any wait seemed to be intolerable in this high-stakes environment where every visit could mean exposure to the virus. (Table 6, column 2, topics 1,5) Patients were critical of visit processes including triage and intake, wait-time, staying in the hallway and rooming activity. The visit processes ranked first compared to fifth and sixth spots pre- and post-COVID-19. Patient privacy was a persistent concern

observed from the topic modeling analysis for both pre and during COVID-19 periods. (word clouds topics 4 and 16, Table 6 column 2, topic 7) This was less of a concern post-COVID-19. Simultaneously, COVID-19 tests (cat-scan, x-ray, and blood tests) appeared frequently in the topics indicating a desire for accurate and but rapid testing.

New topics and concerns that patients reported relevant to the pandemic were identified during COVID-19. Comments on systems issues regarding processes to limit viral spread and concerns over family/visitor restrictions were dominant. Although there was evidence of praise and appreciation of the efforts of staff there was also a high level of scrutiny of the processes encountered during the emergency visit. This scrutiny likely underlines the fears and anxieties related to viral transmission and personal risk. There was also a shift in the type of comments by sentiment analysis classification with an increase in negative sentiments and a decrease in positive sentiments overall.

In the post-COVID period communication issues such as adequate explanations and feeling at ease were dominant, perhaps reflecting the national conversation surrounding

Table 6 Top prevalent topics/subtopics and concerns ordered by weight, by time period

Pre-COVID-19	During COVID-19	Post-COVID-19
technical competence/ knowledgeable	appointment activity	interpersonal manners/ explains
technical competence/ clinical skills	COVID -19/ satisfaction with staff despite stress	interpersonal manners/ put at ease
recommendation	COVID -19/ system issues (staff)	technical competence/ efficient
system issues/ staff	system issues/ appointment access	technical competence / successful treatment
appointment activity	system issues/ appointment wait time	technical competence/ detailed
technical competence/ clinical skills	technical competence/ clinical skills	appointment activity
interpersonal manner/ explain	system issues/ practice environment (privacy)	technical competence/ detailed
negative sentiment	interpersonal manners/ explains	technical competence / successful treatment
technical competence/ successful treatment	technical competence/ knowledgeable	COVID -19/ system issues/ environment concerns

treatment and a high level of scrutiny over best course options and true need for hospitalization versus discharge.

Interestingly, post-first wave of COVID-19, patient comments trended to focus on positive topics like interpersonal manner and technical competence of healthcare providers and staff. (Table 6, column 3 topic 1,2). This positivity is likely a by-product of the goodwill fostered by the professional response of the healthcare team during the height of the pandemic.

7 Conclusion

The analysis demonstrates the shift in patient concerns and perceptions pre, during and post first wave of COVID-19. Sentiment analysis and topic modeling of patient comments draws attention to the issues that became important as the pandemic progressed and receded. Prior to the pandemic, patient comments were largely positive and focused on technical expertise and perceptions of competence which is consistent with findings from studies prior to the pandemic. [6–8] Topic modeling and sentiment analysis can yield critical insights into dynamic patient concerns. The pandemic placed the spotlight on communication, family involvement, infection control, technical expertise, appropriate treatment and perceptions of competence.

Systems issues to address potential viral exposure in waiting and patient care areas became a priority to patients. A clear appreciation of the valiant efforts of healthcare workers was tempered by concerns over systems issues encountered in the emergency visit. These shifting priorities and concerns have implications for monitoring and management of emergency patient comments.

Strategies to address patient concerns around COVID-19 should include scripting and messaging surrounding protocols in place to reduce viral exposure including the necessity of limiting visitors. Frequent updates to patients and family should be maximized. Explanation of treatment and admission triage protocols will help to reduce anxiety surrounding management including need for careful monitoring at home and return precautions.

Although this study is specific to COVID-19, the process of sentiment analysis and topic modeling can help to make sense of the enormous and valuable feedback received from emergency patients that many times goes unexamined. Further work with machine learning to automate this process may yield a useful tool for real time analysis.

Acknowledgements The authors wish to acknowledge Justin Laing for his assistance with the tables and figures.

Authors' contributions SC and SF both contributed to the study design, execution and analysis of the data. HL assisted with coding and data analysis. SC and SF both contributed to the manuscript writing in draft and final form. HL made critical edits and approved the final draft of the manuscript.

Availability of data and material Available upon written request at the discretion of the authors and the Yale IRB with limitations to protect any potentially identifiable information.

Code availability Please see appendix or available upon written request at the discretion of the authors.

Declarations

Ethics approval IRB exemption was obtained for study #2000028733.

Conflicts of interest There are no conflicts of interests or competing interests to declare.

References

- Hefner JL, McAlearney AS, Spatafora N, Moffatt-Bruce SD. Beyond Patient Satisfaction: Optimizing the Patient Experience. *Adv Health Care Manag.* 2019;18.
- Graham B, Endacott R, Smith JE, Latour JM. "They do not care how much you know until they know how much you care": a qualitative meta-synthesis of patient experience in the emergency department. *Emerg Med J.* 2019;36(6):355–63.
- Blackburn J, Ousey K, Goodwin E. Information and communication in the emergency department. *Int Emerg Nurs.* 2019;42:30–5.
- Taylor C, Bengler JR. Patient satisfaction in emergency medicine. *Emerg Med J.* 2004;21(5):528–32.
- López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. *J Gen Intern Med.* 2012;27(6):685–92.
- Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. *J Med Internet Res.* 2013;15(2):e24.
- Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. *J Urol.* 2013;189(6):2269–73.
- Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ open.* 2013;3(1).
- Maramba ID, Davey A, Elliott MN, et al. Web-based textual analysis of free-text patient experience comments from a survey in primary care. *JMIR medical informatics.* 2015;3(2):e20.
- Doing-Harris K, Mowery DL, Daniels C, Chapman WW, Conway M, editors. Understanding patient satisfaction with received healthcare services: a natural language processing approach. AMIA annual symposium proceedings; 2016: American Medical Informatics Association.
- Brody S, Elhadad N, editors. Detecting salient aspects in online reviews of health providers. AMIA Annual Symposium Proceedings; 2010: American Medical Informatics Association.
- Cammel SA, De Vos MS, van Soest D, et al. How to automatically turn patient experience free-text responses into actionable insights: a natural language programming (NLP) approach. *BMC Med Inform Decis Mak.* 2020;20:1–10.
- Shah AM, Yan X, Shah SJ, Khan S, editors. Use of sentiment mining and online NMF for topic modeling through the analysis of patients online unstructured comments. International Conference on Smart Health; 2018: Springer.
- Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. *J Med Internet Res.* 2013;15(11):e239.
- Costello D. Sentiment Analysis: An Emerging Trend That Could Give Hospitals an Edge in Patient Experience 2013 [updated 06/28/2013. Available from: <https://www.beckershospitalreview.com/quality/sentiment-analysis-an-emerging-trend-that-could-give-hospitals-an-edge-in-patient-experience.html>.
- Siegrist Jr R, Madden S. Sentiment analysis turns patients' feelings into actionable data to improve the quality of care. *Sci Emot.* 2011:27–35.
- Hripcsak G HD. Measuring agreement in medical informatics reliability studies. *J Biomed Inform.* 2002;35(2):99–110.
- Ogren PV, Savova GK, Chute CG, editors. Constructing Evaluation Corpora for Automated Clinical Named Entity Recognition. LREC; 2008.
- Ogren PV, Savova G, Buntrock JD, Chute CG, editors. Building and evaluating annotated corpora for medical NLP systems. AMIA Annual Symposium proceedings AMIA Symposium; 2006: American Medical Informatics Association.
- Bird S, editor NLTK: the natural language toolkit. Proceedings of the COLING/ACL on Interactive presentation sessions; 2006: Association for Computational Linguistics.
- Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res.* 2003;3:993–1022.
- Srinivasa-Dekisan B. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy and Keras: Packt Publishing, Ltd.; 2018.
- Rehurek R, & Sojka, P. Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic. 2011;3(2).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sharon Chekijian¹  · Huan Li² · Samah Fodeh³ 

Huan Li
huan.li@yale.edu

Samah Fodeh
samah.fodeh@yale.edu

¹ Department of Emergency Medicine, Yale School of Medicine, CT, New Haven, USA

² Yale School of Public Health, Division of Health Informatics, New Haven, CT, USA

³ Yale School of Public Health, Yale Center for Medical Informatics, Department of Emergency Medicine, Yale School of Medicine, CT, New Haven, USA