Patient Preference and Adherence

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Identifying Patients' Preference During Their Hospital Experience. A Sentiment and Topic Analysis of Patient-Experience Comments via Natural Language Techniques

Jie Yuan[1,*], Xiao Chen[2,*], Chun Yang[3], JianYou Chen[3], PengFei Han[3], YuHong Zhang[2], YuXia Zhang[2]

[1]School of Nursing, Fudan University, Shanghai, 200032, People's Republic of China; [2]Department of Nursing, Zhongshan Hospital of Fudan University, Shanghai, 200032, People's Republic of China; [3]Department of Information, Zhongshan Hospital of Fudan University, Shanghai, 200032, People's Republic of China

*These authors contributed equally to this work

Correspondence: YuXia Zhang; YuHong Zhang, Department of Nursing, Zhongshan Hospital of Fudan University, Room 501, Building 5, Fenglin Road No. 180, Xuhui District, Shanghai, 200032, People's Republic of China, Tel +86 13816881925, Email zhang.yx@aliyun.com; zhang.yuhong@zs-hospital.sh.cn

Background: Open-ended questions in patient experience surveys provide a valuable opportunity for people to express and discuss their authentic opinions. The analysis of free-text comments can add value to quantitative measures by offering information which matters most to patients and by providing detailed descriptions of the service issues that closed-ended items may not cover.

Objective: To extract useful information from large amounts of free-text patient experience comments and to explore differences in patient satisfaction and loyalty between patients who provided negative comments and those who did not.

Methods: We collected free-text comments on a broad, open-ended question in a cross-sectional patient satisfaction survey. We adopted a mixed-methods approach involving a literature review, human annotation, and natural language processing technique to analyze free-text comments. The associations of patient satisfaction and loyalty scores with the occurrence of certain patient comments were tested via logistic regression analysis.

Results: In total, 28054 free-text comments were collected (comment rate: 72.67%). The accuracy of the machine learning approach and the deep learning approach for topic modeling and sentiment analysis was 0.98 and 0.91 respectively, indicating a satisfactory prediction. Participants tended to leave positive comments (69.0%, 19356/28054). There were 22 patient experience themes discussed in the open-ended comments. The regression analysis showed that the occurrence of negative comments about "humanity of care", "information, communication, and education", "sense of responsibility of staff", "technical competence", "responding to requests", and "continuity of care" was significantly associated with a worse patient satisfaction and loyalty, while the occurrence of negative comments about other aspects of healthcare services had no impact on patient satisfaction and loyalty.

Conclusion: The results of this study highlight the interpersonal and functional aspects of care, especially the interpersonal aspects, which are often the "moment of truth" during a service encounter when patients critically evaluate hospital services.

Keywords: patient experience, natural language processing, sentiment analysis, topic modelling, free-text comments

Introduction

Patients, as healthcare recipients, play an essential role in evaluating the quality of care. Gathering, understanding and responding to patients' voices is therefore a popular means of creating a humane healthcare system. Quantitative surveys have been widely adopted to capture patient feedback, and the survey results could serve as a cost-effective method to drive service improvement. For example, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) surveys and Picker Patient Experience Questionnaire-15 (PPE-15) are widely used to measure and improve

Received: 11 March 2025
Accepted: 4 July 2025
Published: 16 July 2025

the quality of hospitalization. However, a major limitation of quantitative surveys is that positive replies are generally given for the closed questions,[1] which leaves little room for quality improvement. Moreover, previous studies have shown that quantitative data provide insufficient detail on the issues that are salient to patients and fail to drive service improvements.[2,3]

To complement quantitative measures, open-ended questions with free-text comments are commonly included in patient experience surveys.[1,3,4] Evidence shows that when patients are presented with both patient narratives and quantitative data, they tend to pay more attention to the narratives.[5] Open-ended questions add value to quantitative measures by offering information that matters most to patients and by providing detailed descriptions of the service issues that closed-ended items may not cover.[2]

Open-ended questions offer the opportunity to obtain substantial actionable information for quality improvement. However, feedback alone is far from sufficient. Free-text comments remain largely unexplored and underutilized,[3] which may be related to the unstructured nature of these replies. Traditionally, extracting meaningful information from raw free-text data requires substantial effort. Cunningham and Wells[3] manually conducted thematic analysis of 6961 free-text comments to identify the proportion of different sentiment and patient experience themes included in the comments. Although manual analysis has produced valuable outcomes, the surge in data volume resulting from information- based feedback mechanisms has given rise to an urgent requirement for scalable methodologies.[4,6]

Natural language processing (NLP) techniques offer promising solutions for efficiently analyzing large free-text datasets. NLP can extract meaningful information and determine the discussion topics that occur in the text. Other industries, such as banks,[7] the tourism industry[8] and marketing,[9] have been quick to embrace this technology to analyze users' needs and preferences. The application of NLP to mine data in patient feedback has also emerged over the last decade.[10] For example, Bovonratwet et al[11] used machine learning-based NLP to conduct sentiment analysis and topic modeling for 1048 patient comments and reported that 25% of comments were negative and 58% were positive, and the negative comments most frequently addressed room conditions and communication. However, analyzing the content of free-text comments remains a nascent technology to the health care industry, and there is little empirical evidence on the relationship between patients' free-text feedback and their overall hospital rating. Moreover, studies on the use of free-text comments to capture patients' needs and preferences have focused mainly on English-speaking users. Little is known about what the Chinese public talks about during their encounters with hospitals.

This study therefore applied a NLP approach to answer the following key questions: What aspects of care do patients discuss? How do patients perceive their hospital journey? And how do commonly expressed patient experience topics, particularly negative comments, correlate with variations in patient satisfaction and loyalty?

Methods

Design

This was a retrospective observational study that used routinely collected patient satisfaction data from June 2022 to June 2023 from a large national medical center in China.

Data Sources

The data were extracted from an electronic system used for collecting patient feedback. One day after discharge, patients were sent a mobile phone text message that included a link to a questionnaire on their care experience during hospitalization. In the text message, patients were informed that their responses would be utilized for analyzing care quality and pursuing a research objective, by completing the questionnaire the participants consented to publication of anonymized responses and direct quotes. If patients were willing to provide feedback, they clicked the embedded link and completed the questionnaire. Three days after the first sent, nonrespondents were sent a reminder. All data were extracted from the database and exported to a text file. The data consisted of the following:

- Patient demographic and diagnostic information, including age, sex, home address, health insurance, primary diagnosis, and length of hospital stay. These data were extracted from the hospital information system.
- Patients' satisfaction and loyalty in relation to the healthcare service. Patients provided feedback on the questions "In general, how satisfied are you with your medical care?", and "To what extent would you recommend your family members and friends to visit this hospital if needed?". The satisfaction score was ranged from 0 to 10, and the recommendation score ranged from 0 to 5.
- Free-text comments (optional) captured patients' feedback on the question "is there any comment you would like to make regarding the service you received?"

In total, 208065 patients were sent the SMS invitation and 38606 responded, with a response rate of 18.54%. Among the respondents, 72.67% (28054/38606) provided free-text comments. These 28054 feedback answers were analyzed, and this sample size was adequately powered to support both qualitative and quantitative analyses.

Data Pre-Processing

The content of the free-text comments was unstructured. For further processing and analysis, we cleaned these comments by removing incorrect punctuation, non-Chinese characters and additional spaces. It is challenging to perform accurate topic modeling for a large corpus, so stop-word removal was used to simplify the dataset.

Qualitative Content Analysis

This study randomly selected 20% of the sample data as learning samples to construct a prediction model. In order to build the model with the best prediction performance, the learning samples were divided into training set, validation set, and testing set with a 60%/20%/20% splitting ratio.[12] The training dataset was used as the learning template to estimate the model parameters. The validation dataset was used to estimate prediction error and avoid overfitting. The test dataset was used to estimate the final model performance. The learning sample was manually coded, while the remaining data would be automatically coded by a machine learning and deep learning approach.

Manual-Coding Approach

To increase the credibility of topic modeling, two researchers with expertise in patient experience independently coded 10% of the total comments from the inpatient survey to develop and refine the coding framework based on the 2012 NHS patient experience framework. The coding framework is shown in Table S1. Each comment was categorized into one or more patient experience themes from the framework. In addition, two researchers used the same dataset to perform sentiment analysis. According to the emotional attributes of the content, the researchers determined and labeled each comment as positive sentiment, neutral sentiment, negative sentiment or mixed sentiment. If the content was a complaint, it was labeled a negative sentiment, while if the content was praise, it was labeled a positive sentiment. If the content was entirely factual, it was labeled neutral (such as "The hospital staff should continue to work hard to create a wonderful service"). A portion of the comments carried two or more sentiments, such as, "When I asked questions, the doctors explained to me carefully and nicely. I like doctors! But the nurses were very impatient, and I felt that they didn't want to spend much time with me". This type of comment was classified as a mixed sentiment.

Manually coded comments were used as the learning template to categorize the remaining comments using machine learning or deep learning algorithms. The interrater agreement for each theme was calculated to limit personal bias. The interrater agreement (Cohen's Kappa) between the two annotators ranged from 0.81 to 0.93, indicating a substantial agreement.[13] During the process of coding, new codes were developed if new content appeared in the comments. Any disagreements were discussed by the team to reach a consensus on the appropriate theme and sentiment of the comment.

Machine Learning Model-Based Coding Approach

The patient experience topic modeling is cast into a multilabel classification problem. We applied six machine learning (ML) approaches to categorize the remaining comments and evaluated the performance of these approaches using the training dataset. The ML approaches included decision tree, support vector machine, logistic regression, XGBoost,

Table 1 The Performance of Sentiment Prediction Models

	Accuracy			Precision			Recall			F-measure		
	Training Set	Valid Set	Test Set	Training Set	Valid Set	Test Set	Training Set	Valid Set	Test Set	Training Set	Valid Set	Test Set
1	0.72	0.74	0.61	0.84	0.84	0.77	0.72	0.74	0.61	0.75	0.76	0.60
2	0.96	0.95	0.69	0.96	0.95	0.71	0.96	0.95	0.69	0.96	0.95	0.69
3	0.96	0.95	0.74	0.96	0.95	0.77	0.96	0.95	0.74	0.96	0.95	0.74
4	0.85	0.83	0.71	0.87	0.86	0.75	0.85	0.83	0.71	0.85	0.84	0.72
5	0.86	0.84	0.74	0.86	0.85	0.76	0.86	0.84	0.74	0.86	0.84	0.74
6	0.96	0.92	0.70	0.96	0.92	0.79	0.96	0.92	0.70	0.96	0.92	0.72
7	0.96	0.95	0.72	0.96	0.95	0.79	0.96	0.95	0.72	0.96	0.95	0.73
8	1.00	0.89	0.91	1.00	0.89	0.91	1.00	0.89	0.91	1.00	0.89	0.91

Notes: 1, Multinomial Naïve Bayes; 2, Decision Tree; 3, Support Vector Machine; 4, XGBoost; 5, Logistic Regression; 6, Random Forest; 7, The integration of Decision Tree, Support Vector Machine and Random Forest; 8, Bert.

multinomial naïve Bayes and random forest. According to the assessment results, the decision tree, support vector machine, and random forest had better performance, with high accuracy, precision, recall, and F-measures (Table 1). To obtain the best-performing model, we used a multiclassifier voting strategy to combine these three high-performing machine learning models to obtain the final classification result in the processing step. As shown in Table 1, the performance metrics of multiclassifier collaborative tagging were excellent. We therefore integrated a decision tree, support vector machine, and random forest using hard voting to construct a classifier to predict patient experience topic. We classified the remaining comments into one or more predefined categories and categorized their sentiment attributes.

Deep Learning Model-Based Coding Approach

Human emotions are complex, and open-closed comments include many mixed sentiments that require contextual analysis. When comments are reviewed manually, contextual information can be accurately analyzed, but this approach is challenging for machine learning method. A new language representation model-BERT, which stands for Bidirectional Encoder Representations from Transformers, has high performance in text-based emotion detection.[14] Therefore, we used a BERT-based model to extract the sentiments in patients' comments. Preclassified data were used as the training set. As shown in Table 2, the performance metrics of the BERT-based model were far better than those of the machine learning models. Furthermore, patient experience comments were classified into five distinct emotion categories, namely happy, angry, sad, surprised, and afraid.

Table 2 The Performance of Patient Experience Themes Prediction Models

	Accuracy			Precision			Recall			F-measure		
	Training Set	Valid Set	Test Set	Training Set	Training Set	Valid Set	Test Set	Training Set	Training Set	Valid Set	Test Set	Training Set
1	0.99	0.99	0.98	0.74	0.74	0.68	0.97	0.96	0.82	0.83	0.83	0.73
2	1.00	1.00	0.98	0.98	0.98	0.74	0.94	0.94	0.78	0.96	0.96	0.76
3	0.98	0.98	0.97	0.64	0.66	0.72	0.89	0.91	0.65	0.73	0.75	0.68
4	0.99	0.99	0.98	0.85	0.86	0.69	0.90	0.90	0.88	0.86	0.86	0.75
5	0.99	0.99	0.99	0.86	0.83	0.69	0.97	0.99	0.91	0.89	0.88	0.76
6	1.00	1.00	0.99	0.96	0.96	0.71	0.94	0.94	0.91	0.95	0.95	0.78
7	1.00	1.00	0.98	0.98	0.98	0.77	0.95	0.96	0.78	0.96	0.97	0.78

Notes: 1, Multinomial Naïve Bayes; 2, Decision Tree; 3, Support Vector Machine; 4, XGBoost; 5, Logistic Regression; 6, Random Forest; 7, The integration of Decision Tree, Support Vector Machine and Random Forest.

 https://doi.org/10.2147/PPA.S526623

Quantitative Analysis

The statistical analysis was conducted using Python and IBM SPSS Statistics 26. To efficiently extract and count each topic, all qualitative data were binarized to address multilabel classification using one-hot encoding. Then, the machine-coded data were imported into SPSS software to describe the characteristics of the discussion topics and sentiments of the comments and to calculate interrater agreement using Cohen's kappa values. To identify which aspects of care patients complained would have an impact on their overall rating of the hospital, we used logistic regression to analyze the relationship between patient satisfaction/loyalty and the occurrence/nonoccurrence of patient experience topics within individual negative patient comments. Because the responses for patient satisfaction and patient loyalty were highly skewed with most scores clustered at the high values, we primarily used the top box scoring method,[15] whereby scores were dichotomized as 5 (maximum score) vs less than 5. Odds ratios were calculated. Independent variables were selected based on evidence from previous studies showing a significant relation to patient experience, such as sex, age, and length of hospital stay. All significance tests were two-sided, and the probability was considered significant when p was < 0.05. No missing data imputation methods were used. Participants who made comments that were with mixed, neutral, or positive sentiments were excluded from the analysis. The Logistic regression was also used to analyze differences in clinical or sociodemographic characteristics between those respondents who made comments, and those respondents who made no comments. We displayed frequently appearing words as "word clouds" to assess the frequency represented by the font size of each word in the comments.

Results

Characteristics of Free-Text Patient Experience Comments

In total, 208065 patients were sent an invitation message; 38606 responded and completed the survey, with an 18.54% response rate. Of those respondents, 72.67% (28054/38606) provided free-text comments. The largest numbers of comments were about nurses (20.15%, 5654/28054), followed by doctors (11.02%, 3092/28054), health care assistants (1.42%, 398/28,054), and nonhealthcare workers (0.10%, 27/28054). Furthermore, 2.17% (609/28054) of the comments that used the term "medical staff" without a particular object. The remaining comments pertained to the environment, medical equipment, or lacked an object.

There were differences in clinical and sociodemographic characteristics between respondents who made comments and those who did not (Table 3). Women, elderly patients, surgical patients, patients without spouses, patients without medical insurance, and patients with lower satisfaction levels and with longer lengths of hospitalization were more likely to comment, while respondents diagnosed with cancer were less likely to comment.

Performance Metrics of Machine Learning Models and Deep Learning Models

Table 2 illustrates the performance metrics of the machine learning models and deep learning models. The accuracy, precision, recall, and F-measures of the integration of the decision tree, support vector machine, and random forest

Table 3 Results of Multivariate Logistic Regression Analysis of Factors Affecting Patients' Behavior of Leaving a Comment

Independent Variables	ß	Standard Error	Wald	OR	95% CI	P value
Women	0.091	0.037	6.099	1.095	1.019~1.177	0.014
Age (per year)	0.009	0.001	49.049	1.009	1.006~1.011	<0.001
Patients without a spouse	0.108	0.044	6.172	1.115	1.023~1.214	0.013
Patients without a medical assurance	0.345	0.052	43.359	1.412	1.274~1.565	<0.001
Surgical patients	0.294	0.040	54.669	1.341	1.241~1.450	<0.001
Patients with a cancer	0.178	0.037	23.114	1.195	1.111~1.285	<0.001
Satisfaction scores	−0.114	0.015	59.441	0.893	0.867~0.919	<0.001
Length of hospitalization	−0.010	0.002	18.702	0.990	0.896~0.995	<0.001

methods for patient experience themes were 0.98, 0.77, 0.78, and 0.78 respectively. For patient experience sentiment, the accuracy, precision, recall, and F-measures for the deep learning models were 0.91.

Sentiment Analysis

Of the 28054 respondents, 69.0% (19356/28054) provided positive comments, 18.0% (5042/28054) provided negative comments, 9.7% (2731/28054) provided neutral comments, and 3.3% (925/28054) provided mixed comments. Positive comments (average 9 words) tended to be shorter, more generic and less detailed than negative comments (average 28 words) and mixed comments (average 47 words). Participants who were older, local or single or not diagnosed with cancer were more likely to leave negative comments (Table S2). Findings from the zero - shot emotion identification indicated that the happy emotion had the highest prevalence, amounting to 48.2% (13522/28054) of the total. Subsequently, the surprised emotion accounted for 16.2% (4544/28054), while the angry, sad, and afraid emotions comprised 15.4% (4321/28054), 13.4% (3759/28054), and 6.8% (1908/28054) respectively.

Patient Experience Themes

Of the 28054 respondents, 16410 provided general comments, such as, "very good" or "very satisfied", while the remaining 11644 commented on certain aspects of care. There were 22 patient experience themes discussed in the open-ended comments (Table 4), and 26.7% (3114/11644) comments discussed more than one theme. Box S1 includes some specific examples of each theme.

As shown in the Table 4, among the respondents who commented on certain aspects of care, the five most commonly mentioned themes were about the "humanity of care" (28.28%, 3293/11644), followed by "information, communication and education" (14.25%, 1659/11644), "food" (13.17%, 1534/11644), "technical competence" (11.04%, 1286/11644), and "ward environment" (10.43%, 1214/11644). The five most common themes in the positive comments were the "humanity of care", "efficacy of treatment", "sense of responsibility of staff", "technical competence", and "food", while the five most common themes in the negative comments were "humanity of care", "information, communication and

Table 4 The Sentiment Distribution of Patient Experience Themes

Patient Experience Topics	Total Comments (n)	Positive Sentiment (n,%)	Negative Sentiment (n,%)	Neutral Sentiment (n,%)	Mixed Sentiment (n,%)
Humanity of care	3293	1624, 49.3%	1099, 33.4%	191, 5.8%	379, 11.5%
Information, communication and education	1659	218, 13.1%	742, 44.7%	572, 34.5%	127, 7.7%
Food	1534	733, 47.8%	516, 33.6%	232, 15.1%	53, 3.5%
Technical competence	1286	942, 73.3%	184, 14.3%	22, 1.7%	138, 10.7%
Ward environment	1214	113, 9.3%	725, 59.7%	281, 23.1%	95, 7.8%
Efficacy of treatment	1186	1146, 96.6%	17, 1.4%	6, 0.5%	17, 1.4%
Sense of responsibility of staff	1154	1035, 89.7%	56, 4.9%	22, 1.9%	41, 3.5%
Post-discharge care	766	354, 46.2%	165, 21.5%	231, 30.2%	16, 2.1%
Access to care	475	6, 1.3%	300, 63.2%	144, 30.2%	25, 5.3%
Equipment	442	17, 3.8%	271, 61.3%	116, 26.2%	38, 8.6%
Efficiency of service process	405	65, 16.0%	203, 50.1%	123, 30.4%	14, 3.5%
Continuity of care	321	91, 28.3%	151, 47.0%	55, 17.1%	24, 7.5%
Responding request	306	87, 28.4%	139, 45.4%	58, 19.0%	22, 7.2%
Standardization of the care procedure	209	76, 36.4%	94, 45.0%	23, 11.0%	16, 7.7%
Insufficient staff	203	21,10.3%	61, 30.0%	116, 57.1%	5, 2.5%
Medical cost	187	9, 4.8%	133, 71.1%	34, 18.2%	11, 5.9%
Involvement of family members	150	3, 2.0%	55, 36.7%	87, 58.0%	5, 3.3%
Privacy	115	16, 13.9%	52, 45.2%	42, 36.5%	5, 4.3
Cross-cultural care	95	15,15.8%	38, 40.0%	35, 36.8%	7, 7.4%
Individualized care	49	12, 24.5%	28, 57.1%	8, 16.3%	1, 2.0%
Excessive treatment	23	1, 4.3%	19, 82.6%	2, 8.7%	1, 4.3%
Physical comfort	10	1, 10.0%	5, 50.0%	4, 40.0%	0, 0.0%

Figure 1 Word clouds of patient experience themes in free-text comments.

education", "ward environment", "food", and "access to care". Word clouds were created to present a visual representation of the text data (Figure 1). Whether in all comments or in positive and negative comments, the proportion of "humanity of care" is the highest. The sentiment distribution of each topic is reported in Table 4.

The Relationship Between the Occurrence of Negative Comments and Patient Loyalty

As shown as Table 5, the regression analysis indicated that the occurrence of negative comments about "humanity of care", "information, communication, and education", "sense of responsibility of staff", "technical competence",

Table 5 The Relationship Between the Occurrence of Negative Comments and Patient Satisfaction and Their Loyalty

Independent Variables	Patient Loyalty				Overall Satisfaction			
	ß	OR	95% CI	P value	ß	OR	95% CI	P value
Humanity of care	−0.858	0.424	0.359–0.501	<0.001	−0.917	0.400	0.340–0.470	<0.001
Information, communication and education	−0.800	0.449	0.367–0.550	<0.001	−0.622	0.537	0.444–0.649	<0.001
Sense of responsibility of staff	−0.747	0.474	0.246–0.912	0.025	−0.785	0.456	0.241–0.863	0.016
Technical competence	−0.725	0.484	0.339–0.691	<0.001	−0.374	0.688	0.497–0.952	0.024
Responding request	−0.703	0.495	0.326–0.752	0.001	−0.813	0.443	0.293–0.670	<0.001
Continuity of care	−0.530	0.589	0.403–0.860	0.006	−0.454	0.635	0.444–0.909	0.013
Standardization of the care procedure	−0.515	0.598	0.375–0.952	0.030	−0.325	0.723	0.468–1.115	0.142
Medical cost	−0.377	0.686	0.470–1.002	0.051	0.044	1.045	0.731–1.495	0.809
Post-discharge care	0.031	1.031	0.718–1.480	0.868	0.125	1.133	0.804–1.597	0.475
Insufficient staff	−0.317	0.728	0.424–1.251	0.250	−0.307	0.736	0.433–1.250	0.257

(Continued)

Table 5 (Continued).

Independent Variables	Patient Loyalty				Overall Satisfaction			
	ß	OR	95% CI	P value	ß	OR	95% CI	P value
Cross-cultural care	−0.142	0.868	0.436–1.727	0.686	−0.333	0.717	0.361–1.422	0.341
Individualized care	−0.308	0.735	0.317–1.706	0.474	0.086	1.090	0.498–2.386	0.829
Involvement of family members	−0.088	0.916	0.522–1.605	0.758	−0.019	0.982	0.567–1.701	0.947
Access to care	−0.002	0.998	0.779–1.280	0.989	0.057	1.059	0.829–1.352	0.647
Efficiency of service process	−0.107	0.898	0.665–1.215	0.487	−0.094	0.910	0.678–1.222	0.531
Efficacy of treatment	0.227	1.255	0.477–3.303	0.645	0.278	1.321	0.504–3.462	0.571
Ward environment	−0.073	0.930	0.782–1.105	0.408	0.084	1.088	0.917–1.291	0.334
Food	0.083	1.087	0.896–1.318	0.400	0.038	1.039	0.857–1.259	0.698
Equipment	0.213	1.238	0.962–1.593	0.098	−0.040	0.961	0.747–1.237	0.759
Physical comfort	−0.110	0.896	0.147–5.470	0.905	−0.290	0.748	0.124–4.512	0.752
Excessive treatment	−0.509	0.601	0.194–1.865	0.378	−0.813	0.443	0.144–1.363	0.156
Privacy	−0.187	0.830	0.470–1.466	0.521	−0.012	0.988	0.567–1.724	0.967

"responding to requests", "continuity of care", and "standardization of the care procedure" was significantly associated with worse patient loyalty (OR = 0.424, 95% CI = 0.359~0.501; OR = 0.449, 95% CI = 0.367~0.550; OR = 0.474, 95% CI = 0.246~0.912; OR = 0.484, 95% CI = 0.339~0.691; OR = 0.495, 95% CI = 0.326~0.752; OR = 0.589, 95% CI = 0.403~0.860; OR = 0.598, 95% CI = 0.375~0.952). The occurrence of negative comments about "humanity of care", "information, communication, and education", "sense of responsibility of staff", "technical competence", "responding to requests", and "continuity of care" was significantly associated with a worse overall satisfaction (OR = 0.400, 95% CI = 0.340~0.470; OR = 0.537, 95% CI = 0.444~0.649; OR = 0.456, 95% CI = 0.241~0.863; OR = 0.688, 95% CI = 0.497~0.952; OR = 0.443, 95% CI = 0.293~0.670; OR = 0.635, 95% CI = 0.444~0.909). The occurrence of negative comments about other patient experience themes had no impact on patient satisfaction and their loyalty.

Discussion

Providing free-text comment boxes enables patients to freely discuss particular aspects of the health care service that are important to them or that have an impact on their overall experience. In this study, 72.67% of the participants provided free-text comments, indicating that patients are active in providing their feedback. Natural language process technology was used to process large amounts of free-text patient experience responses efficiently and to mine meaningful and actionable information for improvement.

Free-text patient experience feedback is unstructured and the texts are always multilabeled, which means that patients discuss more than one topic and that their narrative can be assigned to two or more labels. In this study, 11.10% (3114/28054) of patients commented on two or more specific aspects of care. Therefore, patient experience topic modeling is a multilabel classification task. This type of task is often considered more challenging than single-label text classification.[16] A traditional method for handling the multilabel classification problem is to decompose it into multiple independent binary classification tasks. To address this issue, this study integrated the decision tree, support vector machine, and random forest methods and used a hard voting method to propose multilabel learning algorithms. The machine learning-based multimodel voting ensemble strategy achieved an accuracy of 0.98 for multiclassification tasks. This performance was higher than that of any individual machine learning model, indicating its robust classification performance on the label-imbalanced dataset.[17] Compared with individual models, ensemble models exhibit overall better performance in many other industries.[18]

To address the complex sentiment elements included in the patients' comments, a new language representation model, BERT, which stands for Bidirectional Encoder Representations from Transformers, was used. The BERT approach achieved accuracy of 0.91 for sentiment analysis, indicating its excellent classification performance. Its performance was better than the machine learning model. The sentiment analysis showed that 69.0% (19356/28054) participants provided positive comments, other existing NLP research[11] in healthcare also found the positive emotion accounted for the largest share.

The analysis of the comments revealed that the majority were about nurses and doctors, indicating that interpersonal interactions are patients' main issue of interest during their encounters with hospitals. The positive-to-negative comment ratio was 1:0.26, demonstrating that although most participants experienced positive care, a noteworthy minority reported a negative hospital experience. Moreover, previous studies have shown that negative comments have a greater value for driving changes than positive comments.[19,20] Our study also found that negative comments tended to be longer and more detailed than positive comments and had much richer information. The health care system should monitor the rise of negative voices against services.

Among the patients' discourses included in this content analysis, a wide range of themes were discussed. These themes were divided into interpersonal and functional aspects. The interpersonal aspects included humanity of care, information, education and education, privacy protection, involvement of family members, and responding to requests, which constituted the greatest proportion of patient experience topics. Larson et al[21] also stated that patient experience mainly reflects the interpersonal aspects of health care services. This is especially true for the humanity of care, which has been universally discussed and demonstrated to be the critical attribute of patient experience and satisfaction in previous research.[22,23] This study found that themes associated with positive and negative emotional feedback both focus on the humanity of care (eg, "rude", "friendly"). Rude behavior encountered by patients may trigger dissatisfaction, while friendly behavior may receive praise. Maramba et al[24] conducted a textual analysis of free-text comments from patients and found that "rude" was significantly associated with a worse experience. Wofford et al[25] also reported that patients consistently complained about the interpersonal aspects of care. Improving the interpersonal aspects of care therefore plays a critical role in managing patient experience.

Functional aspects included access to care, food, ward environment, technical competence, efficacy of treatment, physical comfort, error in treatment and coordination of care. Similar findings were presented in previous studies.[26] Many topics appear frequently in traditional hospital-initiated surveys (eg, access to care and physical comfort). There are also some topics that are not typically addressed, such as after-discharge care and coordination of care, indicating that health care organizations are able to use open-ended responses to identify unexpected aspects of care that may not be apparent to hospitals.[9]

Although a wide range of patient experience themes were discussed, this study found that patients had different preferences for specific aspects of care. We analyzed the relationship between the occurrence of negative comments and overall satisfaction and patient loyalty and found that the occurrence of negative comments about "humanity of care", "information, communication, and education", "sense of responsibility of staff", "technical competence", "responding to requests", "continuity of care", and "standardization of the care procedure" was significantly associated with worse patient satisfaction and loyalty, while the occurrence of negative comments about other aspects of care, such as "ward environment", "equipment", and "food" had no impact on patient satisfaction and loyalty. This suggests that different aspects of healthcare service have varying impacts on patient experience. For example, although the ward environment and equipment leave a bad impression on patients, it may not affect their overall evaluation of service quality. On the contrary, when there is a lack of humanistic care in services, insufficient timely and proactive response to needs, or unmet health knowledge needs, patients will have a poor hospital experience. Most of the discussion topics that have a significant impact on patient satisfaction and loyalty pertain to the interpersonal aspects of care. Therefore, this study underscores that the interpersonal aspects of care typically represent the "moment of truth" in a service interaction. When patients make a critical evaluation of these interpersonal aspects of care, they are less likely to recommend their family members or friends to visit the hospital, demonstrating that the interpersonal aspects of care are particularly important to patients. Efforts to enhance interpersonal aspects of care—such as communication skills, empathy training, and care coordination—remain crucial for delivering truly patient-centered care, as these aspects directly influence how patients perceive and engage with their healthcare experiences.

Limitations

This was a single-center study, and our findings therefore may not be generalizable. However, our hospital is a national hospital, and this study is the first to analyze free-text patient experience comments in China. Thus, we suggest that this research provides a starting point for Chinese hospital administrators and clinicians to consider how free-text patient experience comments can assist with health care improvement. In addition, hospital-originated survey may influence the content of patient feedback, while online platforms allow individuals to openly gather, communicate, and share information about their interactions with healthcare services, becoming an essential means of understanding patient experience. Future research should consider the value of online platform. Moreover, the sentiment labelling is a subjective process, and further research should explore other approaches to further validate the findings in this study.

Conclusions

The five most frequent patient experience discussion topics were “humanity of care”, followed by “information, communication and education”, “food”, “technical competence”, and “ward environment”, highlighting the interpersonal and functional aspects of care. The occurrence of negative comments about “humanity of care”, “information, communication, and education”, “sense of responsibility of staff”, “technical competence”, “responding to requests”, and “continuity of care” was significantly associated with worse patient satisfaction and loyalty, demonstrating that the interpersonal aspects of care may hold particular significance for patients.

Abbreviations

HCAHPS, the Hospital Consumer Assessment of Healthcare Providers and Systems; PPE-15, Picker Patient Experience Questionnaire-15; NLP, Natural language processing; ML, machine learning.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable requests.

Ethics Approval and Consent to Participate

The procedure mentioned in this retrospective study involving human participants was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Zhongshan Hospital Fudan University (No. B2020-07R). The need for written informed consent from individual patients was waived by the Ethics Committee of Zhongshan Hospital Fudan University because all the data were anonymized for research purposes.

Acknowledgments

We would like to acknowledge the hard and dedicated work of all the staff in the process of experimental data collection and many valuable ideas put forward in the team discussion.

Author Contributions

Jie YUAN, Xiao CHEN, Chun YANG, YuHong ZHANG and YuXia ZHANG contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jie YUAN, Xiao CHEN, Chun YANG, JianYou CHEN, PengFei HAN, Yuhong ZHANG, and YuXia ZHANG. The first draft of the manuscript was written by Jie YUAN, Xiao CHEN and Chun YANG and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This study has been supported by Shanghai Municipal Health Commission-The Key Discipline of the Three-Year Action Plan for Strengthening the Public Health System Construction in Shanghai (2023–2025).

Disclosure

The authors declare no competing interests in this work.

References

1. Marcinowicz L, Chlabicz S, Grebowski R. Open-ended questions in surveys of patients’ satisfaction with family doctors. *J Health Serv Res Policy.* 2007;12(2):86–89. doi:10.1258/135581907780279639

2. Asprey A, Campbell JL, Newbould J, et al. Challenges to the credibility of patient feedback in primary healthcare settings: a qualitative study. *Br J Gen Pract*. 2013;63(608):e200–e208. doi:10.3399/bjgp13X664252
3. Cunningham M, Wells M. Qualitative analysis of 6961 free-text comments from the first National Cancer Patient Experience Survey in Scotland. *BMJ Open*. 2017;7(6):e015726. doi:10.1136/bmjopen-2016-015726
4. Nawab K, Ramsey G, Schreiber R. Natural language processing to extract meaningful information from patient experience feedback. *Appl Clin Inform*. 2020;11(2):242–252. doi:10.1055/s-0040-1708049
5. Huppertz JW, Otto P. Predicting HCAHPS scores from hospitals' social media pages: a sentiment analysis. *Health Care Manage Rev*. 2018;43 (4):359–367. doi:10.1097/HMR.0000000000000154
6. 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):97. doi:10.1186/s12911-020-1104-5
7. Piris Y, Gay A-C. Customer satisfaction and natural language processing. *J Bus Res*. 2021;124:264–271. doi:10.1016/j.jbusres.2020.11.065
8. Ounacer S, Mhamdi D, Ardchir S, Daif A, Azzouazi M. Customer sentiment analysis in hotel reviews through natural language processing techniques. *Int J Adv Comput Sci Appl*. 2023;14(1). doi:10.14569/IJACSA.2023.0140162
9. Aldunate Á, Maldonado S, Vairetti C, Armelini G. Understanding customer satisfaction via deep learning and natural language processing. *Expert Syst Appl*. 2022;209:118309. doi:10.1016/j.eswa.2022.118309
10. Khanbhai M, Anyadi P, Symons J, Flott K, Darzi A, Mayer E. Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. *BMJ Health Care Inf*. 2021;28(1):e100262. doi:10.1136/bmjhci-2020-100262
11. Bovonratwet P, Shen TS, Islam TW, Ast MP, Haas SB, Su EP, et al. Natural language processing of patient-experience comments after primary total knee arthroplasty. *J Arthroplasty*. 2021;36(3):927–934. doi:10.1016/j.arth.2020.09.055
12. Raykar VC, Saha A. Data split strategiesfor evolving predictive models. In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7–11, 2015, Proceedings, Part I 15*; Springer; 2015:3–19.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174. doi:10.2307/2529310
14. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:181004805*. 2018.
15. Liao L, Chung S, Altamirano J, et al. The association between Asian patient race/ethnicity and lower satisfaction scores. *BMC Health Serv Res*. 2020;20(1):1–11. doi:10.1186/s12913-020-05534-6
16. Du J, Chen Q, Peng Y, Xiang Y, Tao C, Lu Z. ML-Net: multi-label classification of biomedical texts with deep neural networks. *J Am Med Inform Assoc*. 2019;26(11):1279–1285. doi:10.1093/jamia/ocz085
17. Tang P, Yan X, Nan Y, Xiang S, Krammer S, Lasser T. FusionM4Net: a multi-stage multi-modal learning algorithm for multi-label skin lesion classification. *Med Image Anal*. 2022;76:102307. doi:10.1016/j.media.2021.102307
18. Peppes N, Daskalakis E, Alexakis T, Adamopoulou E, Demestichas K. Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0. *Sensors*. 2021;21(22):7475. doi:10.3390/s21227475
19. Riiskjær E, Ammentorp J, Kofoed P-E. The value of open-ended questions in surveys on patient experience: number of comments and perceived usefulness from a hospital perspective. *Int J Qual Health Care*. 2012;24(5):509–516. doi:10.1093/intqhc/mzs039
20. Bjertnaes O, Iversen HH, Skyrud KD, Danielsen K. The value of Facebook in nation-wide hospital quality assessment: a national mixed-methods study in Norway. *BMJ Qual Saf*. 2020;29(3):217–224. doi:10.1136/bmjqs-2019-009456
21. Larson E, Sharma J, Bohren MA. Tunçalp Ö: when the patient is the expert: measuring patient experience and satisfaction with care. *Bull World Health Organ*. 2019;97(8):563. doi:10.2471/BLT.18.225201
22. Ng JH, Luk BH. Patient satisfaction: concept analysis in the healthcare context. *Patient Educ Couns*. 2019;102(4):790–796. doi:10.1016/j.pec.2018.11.013
23. Doing-Harris K, Mowery DL, Daniels C, Chapman WW, Conway M. Understanding patient satisfaction with received healthcare services: a natural language processing approach. In: *AMIA annual symposium proceedings*; American Medical Informatics Association; 2016:524.
24. 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 Med Inform*. 2015;3(2):e20. doi:10.2196/medinform.3783
25. Wofford MM, Wofford JL, Bothra J, Kendrick SB, Smith A, Lichstein PR. Patient complaints about physician behaviors: a qualitative study. *Acad Med*. 2004;79(2):134–138. doi:10.1097/00001888-200402000-00008
26. Chi-Lun-Chiao A, Chehata M, Broeker K, et al. Patients' perceptions with musculoskeletal disorders regarding their experience with healthcare providers and health services: an overview of reviews. *Arch Physiother*. 2020;10(1):1–19. doi:10.1186/s40945-020-00088-6