

Original Paper

Estimating Patient Satisfaction Through a Language Processing Model: Model Development and Evaluation

Shinichi Matsuda^{1*}, PhD; Takumi Ohtomo^{1*}, MSc; Masaru Okuyama²; Hiraku Miyake²; Kotonari Aoki¹, MSc

¹Drug Safety Division, Chugai Pharmaceutical Co Ltd, Tokyo, Japan

²Initiative Inc, Tokyo, Japan

*these authors contributed equally

Corresponding Author:

Shinichi Matsuda, PhD

Drug Safety Division

Chugai Pharmaceutical Co Ltd

2-1-1 Nihonbashi-Muromachi, Chuo-ku

Tokyo, 103-8324

Japan

Phone: 81 8080105061

Email: matsudasni@chugai-pharm.co.jp

Abstract

Background: Measuring patient satisfaction is a crucial aspect of medical care. Advanced natural language processing (NLP) techniques enable the extraction and analysis of high-level insights from textual data; nonetheless, data obtained from patients are often limited.

Objective: This study aimed to create a model that quantifies patient satisfaction based on diverse patient-written textual data.

Methods: We constructed a neural network-based NLP model for this cross-sectional study using the textual content from disease blogs written in Japanese on the Internet between 1994 and 2020. We extracted approximately 20 million sentences from 56,357 patient-authored disease blogs and constructed a model to predict the patient satisfaction index (PSI) using a regression approach. After evaluating the model's effectiveness, PSI was predicted before and after cancer notification to examine the emotional impact of cancer diagnoses on 48 patients with breast cancer.

Results: We assessed the correlation between the predicted and actual PSI values, labeled by humans, using the test set of 169 sentences. The model successfully quantified patient satisfaction by detecting nuances in sentences with excellent effectiveness (Spearman correlation coefficient [ρ]=0.832; root-mean-squared error [RMSE]=0.166; $P<.001$). Furthermore, the PSI was significantly lower in the cancer notification period than in the preceding control period (-0.057 and -0.012 , respectively; 2-tailed $t_{47}=5.392$, $P<.001$), indicating that the model quantifies the psychological and emotional changes associated with the cancer diagnosis notification.

Conclusions: Our model demonstrates the ability to quantify patient dissatisfaction and identify significant emotional changes during the disease course. This approach may also help detect issues in routine medical practice.

(*JMIR Form Res* 2023;7:e48534) doi: [10.2196/48534](https://doi.org/10.2196/48534)

KEYWORDS

breast cancer; internet; machine learning; natural language processing; natural language-processing model; neural network; NLP; patient satisfaction; textual data

Introduction

In any service industry, the goal is to identify and respond to customer needs [1]. In health care, customers are patients, and services must be provided based on whether patients are satisfied with the diagnoses, treatment, and care that they receive. Additionally, in the medical field, patient satisfaction can be

considered in the context of pharmacovigilance (PV). PV is related to monitoring and minimizing the risks of unfavorable events associated with medications, such as adverse drug reactions, and optimizing the benefit-risk profile throughout the drug life cycle [2]. Recently, in the wake of technological data innovations, some traditional PV activities have faced challenges in improving efficiency and quality [3]. To improve PV, it is

essential to consider recruiting patients who have first-hand experience of treatments, but this remains insufficient [4]. Traditional measures, such as spontaneous reporting and electronic health care databases, do not contain information related to patients' emotions; previous research in this area has used questionnaires regarding quality of life (QOL) or patient-reported outcomes [5]. Furthermore, obtaining sufficient and generalizable knowledge is often challenging owing to small sample sizes and limited patient diversity [6].

Insights from patient-derived resources can help improve PV guidelines to better meet patient needs. However, encouraging patient involvement is associated with challenges in obtaining feedback from different patients. Recent PV studies considered data from social networking services (SNSs) [7-9], but research has shown that the benefits of using Twitter and Facebook as patient-derived resources for PV are few [10]. This may be because SNSs contain a significant amount of information unrelated to treatment, making it difficult to extract the necessary data, as the primary scope of these SNSs is not related to the provision of medical care.

Data resources that are primarily focused on collecting patient experiences are rare. An example of a successful method of collecting such information is the web-based patient community operating in the United States, PatientsLikeMe, which has conducted both prospective data collection and evaluation of treatments from the perspective of patients [11]. One conventional method of collecting data regarding patient feedback is through surveys that include questions on QOL. Surveys conducted by PatientsLikeMe use multiple-choice questionnaires to ask patients about their disease complaints [12]. However, as most QOL questions are developed from the perspective of the health care professional, they tend to focus only on the intentions of those asking the questions. A recent study showed that patients with COVID-19 tended to describe broad aspects of care that mattered to them in the comment field, regardless of the focus of the survey question [13]. This highlights the importance of patient-written narratives, as essential opinions of patients may be overlooked if only prespecified questionnaires are used. Another key challenge is obtaining data that reflect the opinions of a wide range of patients receiving various treatments.

Regarding such data, some patients in Japan who had adopted the custom of writing *tōbyōki*, which are diaries about their longitudinal experience with diseases [14], began posting their *tōbyōki* as blogs in the mid-1990s. These *tōbyōki* blogs, in combination with natural language processing (NLP) [15], facilitate a qualitative understanding of treatment experiences and feelings [16,17]. Nonetheless, a qualitative description alone is not sufficient to make decisions that could improve patient care, and effective methods for visualizing patient anxieties and frustrations are needed.

Based on recent research trends, we believe the following 3 research issues should be addressed:

1. Critical information is missed when only computers are used to classify results based on a preset group of options. Thus, the contents need to be reviewed manually. This takes considerable time before starting a discussion for action.

2. Analyzing qualitative data quantitatively to investigate the significance of differences is challenging. For example, sentiment analysis programs can identify a particular sentiment category as “negative” for a specific group of texts based on machine learning; however, this classification alone does not indicate the intensity of that negativity. Although some studies have quantitatively shown the degree of sentiment, the validity of such estimations in the health care domain has been limited [18]. Hence, manual review by humans is required to prioritize issues for action. Thus, a system that can quantitatively assess the impact of qualitative content in a timely fashion is needed.
3. When using supervised machine learning to classify groups according to patients' comments, researchers must specify the classification groups beforehand. If a patient has a complaint that does not fit into the prespecified classification groups, it is challenging to capture the complaint using previous classification methods. Thus, we believe that these types of classification approaches have led to the loss of a significant percentage of social media data, which is unfortunate because patients' comments on social media can yield novel, unanticipated complaints given that they do not have the same restrictions that they would have with structured questionnaires [19].

We believe that the quantitative assessment of patient satisfaction using data from the narratives of diverse patient populations is of significant value; however, no such efforts have been reported. If an efficient tool to obtain patient feedback from written texts could be developed, it could be widely applied to improve health care services. In this study, we aimed to create an NLP model that quantifies patient satisfaction based on diverse patient-written textual data.

Methods

Description and Processing of Data

We have collected anonymous, publicly available data from *tōbyōki* blogs written in Japanese from the internet [20]. Individual *tōbyōki* blogs were manually tagged by disease based on the blog title or introduction page. Between 1994 and 2020, the most frequently reported disease on *tōbyōki* blogs was breast cancer (n=6669 blog entries), followed by depression (n=3295), cervical cancer (n=1231), and rheumatoid arthritis (n=1144). We focused on breast cancer because it has the largest number of entries and is the most common cancer in women worldwide [21]. Additionally, as the recurrence of breast cancer exerts a severe psychological distress on patients [22], there is a substantial need to evaluate patient satisfaction during treatment.

Overall Flow of Model Construction and Evaluation

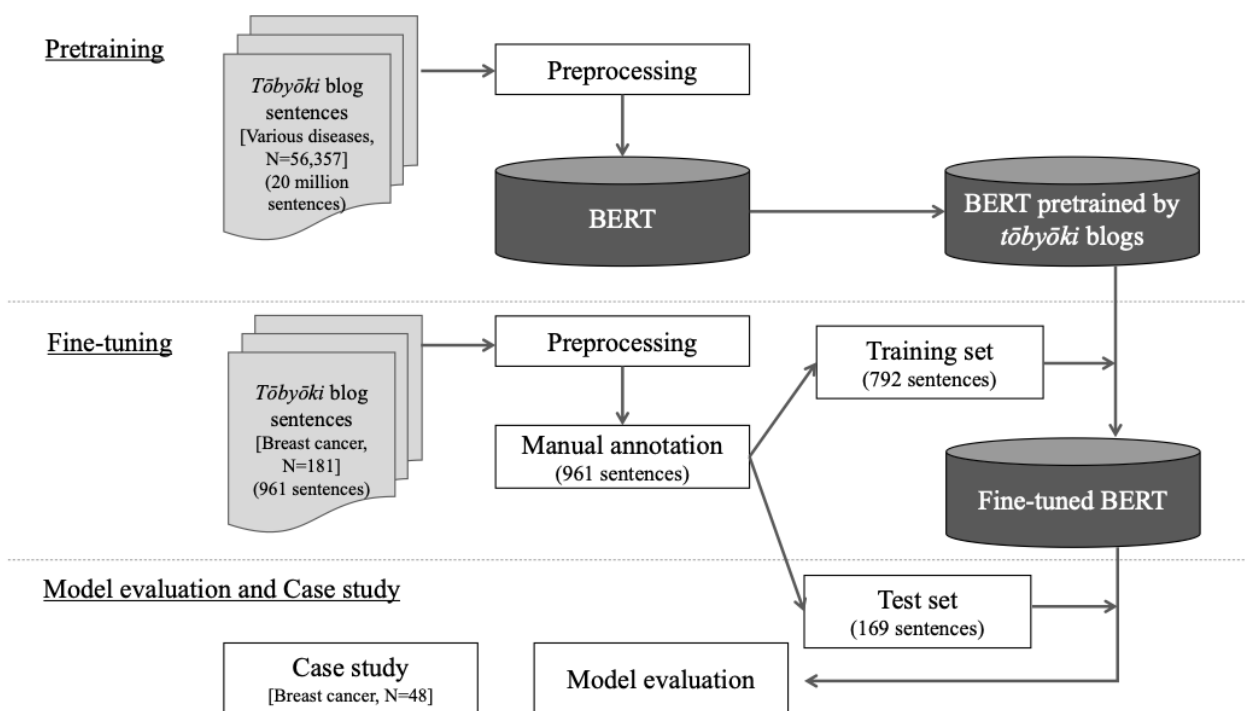
A systematic review revealed no unified definition of patient satisfaction [23]. It is assumed that patient satisfaction is the result of a combination of treatment efficacy, quality of care, and QOL. Here, the patient satisfaction index (PSI) was defined as a numerical value ranging from -1.0 to 1.0, representing the most dissatisfied and satisfied states, respectively. We applied NLP and machine learning (ML) to develop a quantitative method for evaluating PSIs using textual information. Currently, most approaches to text analysis are based on sentiment analysis,

which uses a simple sum of each word's emotional score [24]. However, this approach limits the ability to capture emotional subtleties. Another critical issue when dealing with words is how to consider the context of each sentence. A previous study reported that the performance of ML models for textual data is inadequate in the absence of contextual considerations [25]. To address these issues, we applied a neural network-based model known as bidirectional encoder representation from transformers (BERT) [26]. It uses word embedding, which helps capture sentence nuance more naturally than sentiment analysis [24]. In addition, BERT provides improved performance because it incorporates the context before and after the words to capture nuances better than other models.

We attempted to predict PSIs using a regression approach. The overall flow of the model construction is shown in Figure 1, and the data flowchart is shown in Figure S1 in Multimedia Appendix 1. For the training process, we collected 20 million *tōbyōki* blog sentences extracted from disease blogs written by 56,357 patients concerning 1402 distinct diseases, and then the preprocessing was performed to remove the extraneous information (Table S1 in Multimedia Appendix 1) [27]. In

general, pretraining requires a vast amount of data and tends to be highly computational and time-consuming. To complete this process efficiently, we implemented transfer learning using the publicly available Japanese version of BERT [28]. We then prepared the training data for fine-tuning, which enabled BERT to predict PSIs. Specifically, we prepared 961 sentences extracted randomly from *tōbyōki* blog entries written by 181 patients with breast cancer. Next, 3 reviewers (SM, TO, and HM) independently reviewed each sentence and provided a numerical value for the PSI as the actual label (PSI labeling guidelines are presented in Table S2 in Multimedia Appendix 1; statistics for annotation results are presented in Table S3 in Multimedia Appendix 1); the mean value of the reviewers' PSI labels was treated as the actual PSI for each sentence to ensure the model's robustness by avoiding possible effects of differences between the 3 labels. Subsequently, we randomly divided these sentence data into a training data set (792 sentences) and an unseen test data set (169 sentences) to build the model and evaluate its effectiveness (the characteristics of the data for fine-tuning are presented in Table S4 in Multimedia Appendix 1).

Figure 1. Flow diagram of model construction. The overall flow of model construction consists of a pretraining phase, fine-tuning phase, and model evaluation phase. During pretraining, the bidirectional encoder representation from transformers (BERT) model obtains the spread of words, a mechanism also known as word embedding, from a large data set (in this case, *tōbyōki* blogs). Next, fine-tuning, a supervised learning process, replaces the output layer of the model for a specific task (in this case, predicting patient satisfaction index [PSI] from the texts).



Using the test data set, we conducted a sentiment analysis based on the Japanese sentiment polarity dictionary [29]. Researchers at Tohoku University, a national university in Japan, created this dictionary by collecting nouns and declinable words in Japanese and manually adding polarity information ranging from -1.00 (negative) to 1.00 (positive) [30]. Next, using the same test set, we evaluated the BERT model's performance based on the correlation between the predicted and actual PSI.

Change in PSI After Cancer Notification

To confirm the reliability of the prepared BERT model in evaluating the PSI for actual cases, we conducted an evaluation by predicting the PSI before and after cancer notification. The "cancer notification period" was defined as the time when any blog entry containing the word "cancer notification (in Japanese)" was posted; each of these entries was manually confirmed by a reviewer. In contrast, the "control period" was defined as the time covered by 10 entries posted before the 120 days preceding the cancer notification period (Figure S2 in

[Multimedia Appendix 1](#)). We selected 48 blogs with both a blog entry for the cancer notification period and the preceding control period and computed the PSI in each period; the detailed procedure is presented in Table S5 in [Multimedia Appendix 1](#).

Statistical Analyses

To evaluate the model's capacity, the Spearman correlation coefficient (ρ) and the root-mean-squared error (RMSE) were used. A paired, 2-tailed t test was conducted to compare the PSIs between the cancer notification and control periods. To compare BERT models with and without pretraining, we used a statistical comparison test for 2 correlations [31]. All statistical analyses were performed using R (version 3.6.2; The R Foundation for Statistical Computing), with $P < .05$ being considered statistically significant.

Ethical Considerations

We collected anonymous, publicly available data from web-based sources. Per the copyright law issued by the Japanese Agency for Cultural Affairs (Article 47-7), we adhered to regulations allowing the reproduction and adaptation of copyrighted works onto recording media for information analysis within specified limits [32]. To ensure the privacy of individuals, the breast cancer blog data used in this study were meticulously reviewed. They were visually confirmed to be free of personally identifiable information such as the blog authors' names, handle names, and dates of birth. This study did not necessitate institutional review board approval, aligning with the Ethical Guidelines for Life Sciences and Medical Research Involving Human Subjects [33].

Results

Evaluation of the BERT Model

As part of the dictionary-based approach, we conducted sentiment analysis. The correlation between the actual PSI labeled by humans and the predicted sentiment score was very low ($\rho = 0.127$; $RMSE = 0.563$; $P = .10$; Figure S3 in [Multimedia Appendix 1](#)). This result shows that sentiment analysis based on the widely used Japanese sentiment polarity dictionary is not appropriate for predicting the PSI.

Using the same test set, we evaluated the correlation between the predicted and actual PSI. The BERT model achieved excellent effectiveness as an NLP model for predicting PSIs ($\rho = 0.832$; $RMSE = 0.166$; $P < .001$; Figure 2). PSI prediction using the BERT model without pretraining was less reliable than that using the model with pretraining ($\rho = 0.554$; $RMSE = 0.250$; $P < .001$; Figure S4 in [Multimedia Appendix 1](#)). Furthermore, a statistical comparison test for 2 correlations demonstrated that they were significantly different ($P < .001$). This reaffirms that the model's performance is the highest in the BERT model with pretraining.

We reviewed the sentences with the highest and lowest PSIs (Table 1) and found that our model was able to evaluate the nuances of each sentence. For example, a sentence expressing appreciation for help from a patient's family members and friends yielded a high predicted PSI of 0.45. In contrast, a situation where patients believed that their condition was improving despite experiencing adverse drug reactions yielded a modest predicted PSI of 0.18. Similarly, a sentence describing considerable hardships from radiotherapy yielded the lowest PSI, which was -0.58 .

Figure 2. Correlation between the predicted and actual patient satisfaction index (PSI). Predicted versus actual PSI using the test set.

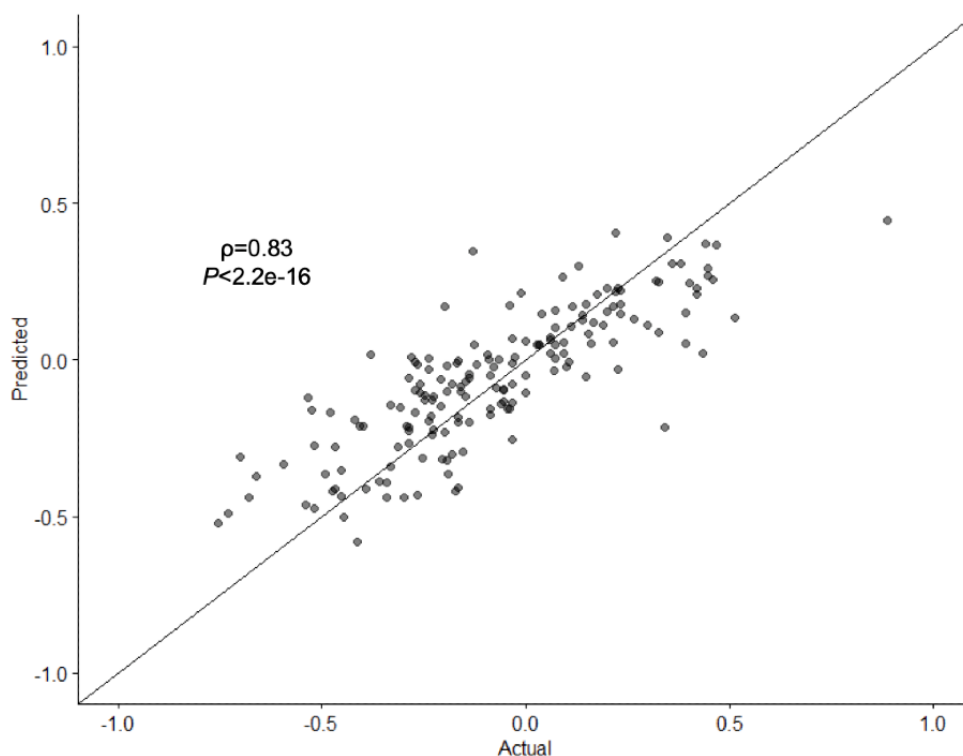


Table 1. Predicted patient satisfaction index (PSI) corresponding to sentences in *tōbyōki* blogs.

Rank	PSI sentence examples ^a	PSI
1	Blindsided by the notification of my illness, I found immense joy and amazement in being able to accomplish so many things in just 6 months, with all of my friends waiting for me, bolstering my happiness.	0.45
2	Perhaps the new anti-nausea medication is working better than expected.	0.41
17	The results of the echocardiogram showed no abnormalities at all.	0.23
25	The only reason I can manage to tolerate the side effects is that I believe my condition is becoming better.	0.18
105	I told the doctor I wanted a chest scan, but he said he would not do it until I had more symptoms.	-0.11
149	I really do not like tests, and no matter how many times I have them, I never get used to them.	-0.33
171	I have had a number of severe depressive episodes, and I have also suffered from menopausal symptoms caused by hormone therapy.	-0.52
172	The radiation treatment was quite painful.	-0.58

^aThese are translations of original sentences written in Japanese. Some minor changes have been made to protect patient privacy. Ranks represent the results ordered from the highest to the lowest predicted PSI for the test set.

Changes in PSI After Cancer Notification

To confirm the applicability of this study’s BERT model in assessing the PSI for actual cases, we compared the PSI distribution between the time of cancer notification and the time before cancer notification using 48 *tōbyōki* blogs written by patients with breast cancer. Many sentences tended to have neutral PSIs (the proportions of PSIs between -0.2 and 0.2 were 1725/2097, 82.3% in the cancer notification period and 7330/8983, 81.6% in the control period; Figure 3). The proportion of sentences showing a negative PSI was higher in the cancer notification period (1342/2097, 64%) than in the control period (4497/8983, 50.1%). In addition, the negative PSI distribution in each sentence was slightly higher in the cancer notification period (mode -0.05) than in the control period (mode 0.05). These results indicate that many sentences in *tōbyōki* blogs had a neutral tone, even when they were written about cancer notification. Overall, the PSI was lower in the cancer notification period than in the control period.

To examine the aforementioned differences in detail, we compared the mean PSI values at each period. Although the SD of the PSI distribution was similar in the cancer notification

(SD 0.042) and control (SD 0.045) periods, the mean PSI was significantly lower in the cancer notification period than in the control period (-0.057 and -0.012, respectively; $t_{47}=-5.392$; $P<.001$; Figure 4A). This result suggests that cancer notification adversely affects the PSI, as expected. However, the average negative effect associated with cancer notification may have been partly diminished owing to the many neutral sentences in the cancer notification and control periods, as shown in Figure 3. This suggests that some sentences expressing shock concerning the notification may be overlooked when relying on the overall mean PSI alone. Hence, we focused on the sentences with the lowest PSI values, which were expected to highlight the most negative effects associated with cancer notification. The mean difference in the PSI between the cancer notification and control periods became apparent when the bottom 5 sentences were compared (-0.298 and -0.144, respectively; $t_{47}=7.214$; $P<.001$; Figure 4B). Furthermore, the SD of the PSI distribution was larger during the cancer notification period (SD 0.128) than during the control period (SD 0.090). These results suggest that focusing on the lower end of the PSI successfully highlights the negative effects of cancer notification.

Figure 3. Distribution of the predicted patient satisfaction index (PSI) before and after cancer notification. Distribution of the predicted PSI at the time of and before cancer notification (defined as the control period). The figure shows the predicted PSI for each sentence in each period.

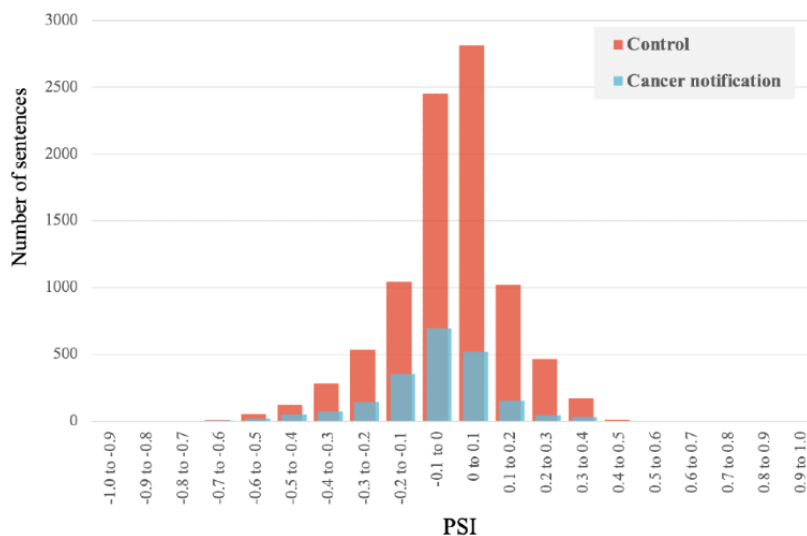
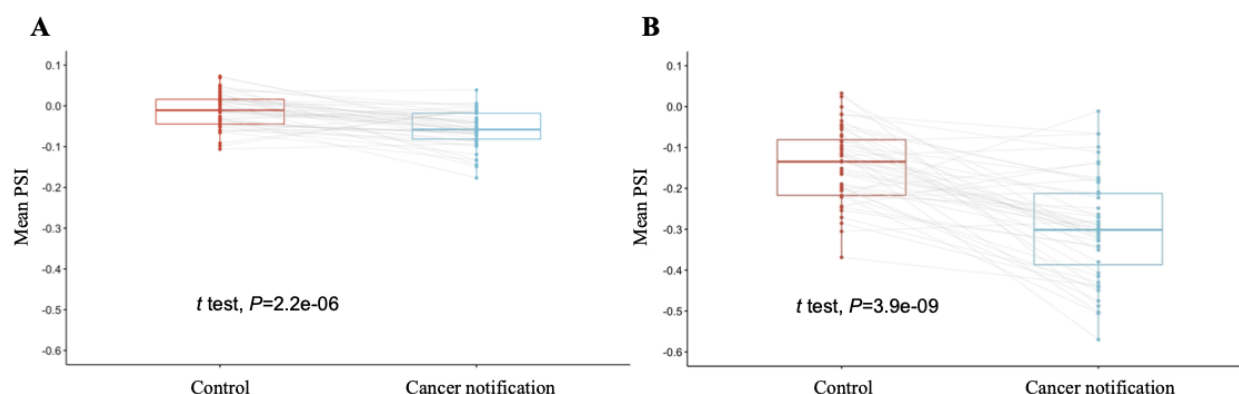


Figure 4. Patient satisfaction index (PSI) comparison before and after cancer notification. (A) Comparison of the mean PSI in the cancer notification and control periods without limiting. (B) Comparison of the mean PSI in the cancer notification and control periods limited to the 5 sentences with the lowest PSI values.



Discussion

Principal Findings

This study showed the successful development of a model that quantifies patient satisfaction based on patient-written textual data. The most recent studies on patient satisfaction have used relatively small samples [34,35]. However, our approach, which involved the use of an ML model with a vast amount of patient-written text from the internet, presented a realistic method for estimating patient satisfaction that can help improve health care services.

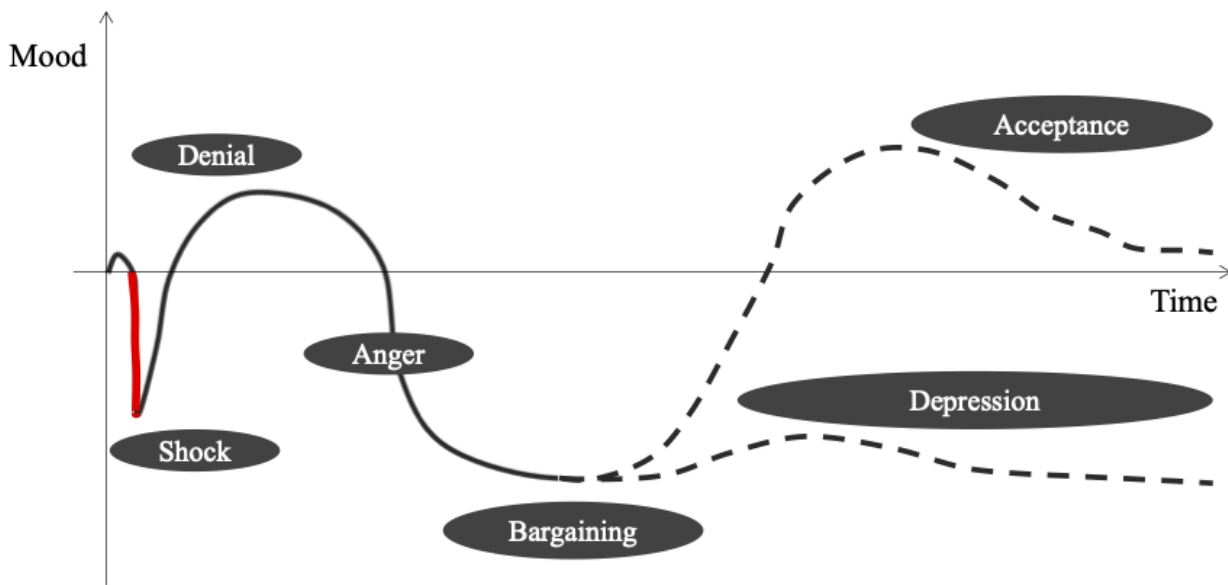
When we applied the neural network-based BERT model, the predicted PSI was strongly correlated with that judged by human interpretation ($\rho=0.832$; Figure 2), and the model became more reliable with pretraining ($\rho=0.554$; Figure S4 in Multimedia Appendix 1). This indicates the importance of pretraining specific to patients' experiences in PSI prediction; therefore, pretraining using a large amount of data is essential to achieve high accuracy. Although a recent systematic review showed that there are many studies on NLP and ML using patient-written free texts [36], few studies have attempted text-based quantification, and no previous studies are available that could provide a benchmark for comparison with our model. Nevertheless, in 1 study on hospital care that evaluated the correlation between scores obtained from a patient questionnaire and those derived from sentiment analysis of free text comments, Spearman correlation coefficient did not exceed 0.50 [37], indicating the difficulty in achieving high accuracy. These findings also support our model's excellent effectiveness in such a situation. Moreover, when we reviewed the sentences paired with PSIs (Table 1), we found that the model's predictions considered the degree of meaning (the nuances of the language expressions) in the sentences, which is difficult to predict using conventional sentiment analysis. Collectively, these findings suggest that this model, which was trained on an immensely high volume of data from *tōbyōki* blogs, can be practically applied to determine accurate PSIs.

We validated the model's applicability in assessing emotional responses to cancer notification, which usually occurs in the

early phase of patients' cancer journey. As expected, PSI was significantly lower during the cancer notification period than during the preceding control period (Figure 4), indicating that the model can quantify the psychological and emotional changes associated with the notification. When limited to the 5 lowest-ranking sentences, the difference in PSIs between the cancer notification and control periods increased, suggesting that the amplitude of negative emotions is high for some people, as reflected by the larger SD in the cancer notification period (SD 0.128) than in the control period (SD 0.090). Understanding patient satisfaction at each stage of treatment during routine clinical practice is essential for improving the medical care system in a patient-centered fashion because there may be discrepancies in the disease burden recognized by patients and clinicians [38]. Moreover, this result is consistent with Elisabeth Kübler-Ross's 5-stage model of death and dying (Figure 5) [39]. In the Kübler-Ross model, a cancer notification induces a "shock," represented as the steepest negative impact (shown as a thick red line). Although several factors, including both personal (eg, age, sex, and religion) and country (eg, country's health policies and medical environment) levels, affect patient satisfaction, it is reasonable to assume that cancer notification is universally considered to be a negative event that causes a decline in patient satisfaction regardless of the differences in these factors. To our knowledge, our model succeeded in numerically expressing this phenomenon for the first time.

Our method of quantitatively identifying cancer-related anxieties can greatly contribute to future NLP research by highlighting patients' medical needs efficiently. Although research using qualitative approaches, such as content analysis, is helpful in analyzing feedback from patients, qualitative review tends to be time-consuming and is associated with many challenges in analyzing information. Our approach, which is based on supervised ML using BERT, can significantly reduce the cost of human review. While many studies have reported achieving good results by applying BERT to both classification [40,41] and regression tasks [42], this study brings a unique perspective by demonstrating the potential usefulness of BERT in predicting patient satisfaction using a regression approach on patient-written textual data.

Figure 5. The 5-stage model of death and dying by Elisabeth Kübler-Ross. The thick red line denotes the “shock” induced by cancer notification. The dotted line indicates a fork in the path between acceptance and depression.



The current findings suggest that our model can be applied as a quantitative index of patient satisfaction not only in the field of PV but also in the field of health care services, where it can be used to determine the effectiveness of communication with health service providers and other related factors [43]. Considering that many hospitals in the United States continue to estimate patient satisfaction with treatment using questionnaires or Twitter [44,45], our approach represents a major breakthrough. Patients may be dissatisfied and anxious about services, including treatments, in their disease-fighting experience. In addition, this model can be applied to diverse areas of research subjects in social psychology, where quantitative consideration of population-level patient emotions may play an essential role in advancing knowledge in the field.

Limitations

This study has several limitations. First, when applying our method cross-culturally, it is necessary to consider that the writers' thoughts, culture, beliefs, and customs affect their content. Second, in our model, we exclusively used breast cancer data; therefore, the model may not be generalizable to other diseases, especially those that affect male patients. Therefore, as a scope for future work, data relating to both male and female patients should be analyzed using blogs on other diseases. Third, patient satisfaction inherently contains some level of subjectivity. It encompasses multiple facets, such as treatment outcomes, communication experience with health care providers,

patients' understanding of their disease, and the overall patient experience. Although this study focuses on an overall assessment of their disease condition and treatment reflected through their blog entries, our approach does not necessarily capture all aspects of patient satisfaction. Fourth, the subjectivity inherent in manual labeling procedures and the limitation of having a relatively small number of manual labels represent another constraint. Although we took careful measures to reduce potential bias and inconsistency, this aspect is subject to individual interpretation and understanding. Finally, we recognize the limitation in our definition of the “cancer notification period.” In our approach, 1 of our researchers manually read and verified the relevance of the content of the selected blogs to cancer notification. While this manual review ensured that the selected blogs were indeed about cancer notification, there may exist instances of cancer notification that use different terminology. Consequently, there may be missed cases where cancer notification is described using alternative expressions.

Conclusions

We have proposed a distinctive approach for estimating patient satisfaction using patient-written textual data. Visualizing the patient's journey and determining the causes of varying patient satisfaction will identify problems in routine medical practice and provide services to resolve these problems.

Acknowledgments

The authors would like to thank Matthew McKeehan for his assistance with drafting the manuscript. We would like to thank Editage for editing and reviewing this manuscript in English. This study received no funding.

Conflicts of Interest

SM, TO, and KA work with Chugai Pharmaceutical Co, Ltd. MO and HM work with Initiative Inc. All other authors declare that they have no competing interests.

Multimedia Appendix 1

eTables and eFigures.

[\[DOCX File , 447 KB-Multimedia Appendix 1\]](#)

References

1. Lovelock CH, Patterson P, Wirtz J. *Services Marketing: An Asia-Pacific and Australian Perspective*, 6th ed. Melbourne: Pearson Australia; 2015.
2. Fornasier G, Francescon S, Leone R, Baldo P. An historical overview over pharmacovigilance. *Int J Clin Pharm* 2018;40(4):744-747 [FREE Full text] [doi: [10.1007/s11096-018-0657-1](https://doi.org/10.1007/s11096-018-0657-1)] [Medline: [29948743](https://pubmed.ncbi.nlm.nih.gov/29948743/)]
3. Edwards IR. The future of pharmacovigilance: a personal view. *Eur J Clin Pharmacol* 2008;64(2):173-181 [doi: [10.1007/s00228-007-0435-9](https://doi.org/10.1007/s00228-007-0435-9)] [Medline: [18172624](https://pubmed.ncbi.nlm.nih.gov/18172624/)]
4. Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med* 2010;362(10):865-869 [FREE Full text] [doi: [10.1056/NEJMp0911494](https://doi.org/10.1056/NEJMp0911494)] [Medline: [20220181](https://pubmed.ncbi.nlm.nih.gov/20220181/)]
5. Jefford M, Ward AC, Lisy K, Lacey K, Emery JD, Glaser AW, et al. Patient-reported outcomes in cancer survivors: a population-wide cross-sectional study. *Support Care Cancer* 2017;25(10):3171-3179 [doi: [10.1007/s00520-017-3725-5](https://doi.org/10.1007/s00520-017-3725-5)] [Medline: [28434095](https://pubmed.ncbi.nlm.nih.gov/28434095/)]
6. Poole KG. Patient-experience data and bias: what ratings don't tell us. *N Engl J Med* 2019;380(9):801-803 [FREE Full text] [doi: [10.1056/NEJMp1813418](https://doi.org/10.1056/NEJMp1813418)] [Medline: [30811905](https://pubmed.ncbi.nlm.nih.gov/30811905/)]
7. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts. *Drug Saf* 2017;40(4):317-331 [FREE Full text] [doi: [10.1007/s40264-016-0491-0](https://doi.org/10.1007/s40264-016-0491-0)] [Medline: [28044249](https://pubmed.ncbi.nlm.nih.gov/28044249/)]
8. Convertino I, Ferraro S, Blandizzi C, Tuccori M. The usefulness of listening social media for pharmacovigilance purposes: a systematic review. *Expert Opin Drug Saf* 2018;17(11):1081-1093 [doi: [10.1080/14740338.2018.1531847](https://doi.org/10.1080/14740338.2018.1531847)] [Medline: [30285501](https://pubmed.ncbi.nlm.nih.gov/30285501/)]
9. Tricco AC, Zarin W, Lillie E, Jeeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. *BMC Med Inform Decis Mak* 2018;18(1):38 [FREE Full text] [doi: [10.1186/s12911-018-0621-y](https://doi.org/10.1186/s12911-018-0621-y)] [Medline: [29898743](https://pubmed.ncbi.nlm.nih.gov/29898743/)]
10. Caster O, Dietrich J, Kürzinger ML, Lerch M, Maskell S, Norén GN, et al. Assessment of the utility of social media for broad-ranging statistical signal detection in pharmacovigilance: results from the WEB-RADR project. *Drug Saf* 2018;41(12):1355-1369 [FREE Full text] [doi: [10.1007/s40264-018-0699-2](https://doi.org/10.1007/s40264-018-0699-2)] [Medline: [30043385](https://pubmed.ncbi.nlm.nih.gov/30043385/)]
11. Brownstein CA, Brownstein JS, Williams DS, Wicks P, Heywood JA. The power of social networking in medicine. *Nat Biotechnol* 2009;27(10):888-890 [doi: [10.1038/nbt1009-888](https://doi.org/10.1038/nbt1009-888)] [Medline: [19816437](https://pubmed.ncbi.nlm.nih.gov/19816437/)]
12. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol* 2011;29(5):411-414 [FREE Full text] [doi: [10.1038/nbt.1837](https://doi.org/10.1038/nbt.1837)] [Medline: [21516084](https://pubmed.ncbi.nlm.nih.gov/21516084/)]
13. Guney S, Daniels C, Childers Z. Using AI to understand the patient voice during the Covid-19 pandemic. *NEJM Catal Innov Care Deliv* 2020;1(2) [FREE Full text]
14. Kadobayashi M. *As a Source of Power to Live: Sociology of Tōbyōki for Cancer* [in Japanese]. Tokyo: Seikaisha Press; 2011.
15. Hirschberg J, Manning CD. Advances in natural language processing. *Science* 2015;349(6245):261-266 [doi: [10.1126/science.aaa8685](https://doi.org/10.1126/science.aaa8685)] [Medline: [26185244](https://pubmed.ncbi.nlm.nih.gov/26185244/)]
16. Matsuda S, Aoki K, Tomizawa S, Sone M, Tanaka R, Kuriki H, et al. Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. *JMIR Public Health Surveill* 2017;3(1):e10 [FREE Full text] [doi: [10.2196/publichealth.6872](https://doi.org/10.2196/publichealth.6872)] [Medline: [28235749](https://pubmed.ncbi.nlm.nih.gov/28235749/)]
17. Matsuda S, Ohtomo T, Tomizawa S, Miyano Y, Mogi M, Kuriki H, et al. Incorporating unstructured patient narratives and health insurance claims data in pharmacovigilance: natural language processing analysis of patient-generated texts about systemic lupus erythematosus. *JMIR Public Health Surveill* 2021;7(6):e29238 [FREE Full text] [doi: [10.2196/29238](https://doi.org/10.2196/29238)] [Medline: [34255719](https://pubmed.ncbi.nlm.nih.gov/34255719/)]
18. He L, Yin T, Hu Z, Chen Y, Hanauer DA, Zheng K. Developing a standardized protocol for computational sentiment analysis research using health-related social media data. *J Am Med Inform Assoc* 2021;28(6):1125-1134 [FREE Full text] [doi: [10.1093/jamia/ocaa298](https://doi.org/10.1093/jamia/ocaa298)] [Medline: [33355353](https://pubmed.ncbi.nlm.nih.gov/33355353/)]
19. Rozenblum R, Greaves F, Bates DW. The role of social media around patient experience and engagement. *BMJ Qual Saf* 2017;26(10):845-848 [doi: [10.1136/bmjqs-2017-006457](https://doi.org/10.1136/bmjqs-2017-006457)] [Medline: [28428244](https://pubmed.ncbi.nlm.nih.gov/28428244/)]
20. Tōbyōki blog collection [in Japanese]. TOBYO. URL: <https://www.tobyoy.jp/> [accessed 2023-04-11]
21. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68(6):394-424 [FREE Full text] [doi: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492)] [Medline: [30207593](https://pubmed.ncbi.nlm.nih.gov/30207593/)]

22. Skandarajah AR, Lisy K, Ward A, Bishop J, Lacey K, Mann B, et al. Patient-reported outcomes in survivors of breast cancer one, three, and five years post-diagnosis: a cancer registry-based feasibility study. *Qual Life Res* 2021;30(2):385-394 [doi: [10.1007/s11136-020-02652-w](https://doi.org/10.1007/s11136-020-02652-w)] [Medline: [32997334](https://pubmed.ncbi.nlm.nih.gov/32997334/)]
23. Batbaatar E, Dorjdagva J, Luvsannyam A, Amenta P. Conceptualisation of patient satisfaction: a systematic narrative literature review. *Perspect Public Health* 2015;135(5):243-250 [doi: [10.1177/1757913915594196](https://doi.org/10.1177/1757913915594196)] [Medline: [26187638](https://pubmed.ncbi.nlm.nih.gov/26187638/)]
24. Denecke K, Deng Y. Sentiment analysis in medical settings: new opportunities and challenges. *Artif Intell Med* 2015;64(1):17-27 [doi: [10.1016/j.artmed.2015.03.006](https://doi.org/10.1016/j.artmed.2015.03.006)] [Medline: [25982909](https://pubmed.ncbi.nlm.nih.gov/25982909/)]
25. 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 [FREE Full text] [doi: [10.2196/jmir.2721](https://doi.org/10.2196/jmir.2721)] [Medline: [24184993](https://pubmed.ncbi.nlm.nih.gov/24184993/)]
26. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv. Preprint posted online on May 24 2019 2019 [doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805)]
27. Morita H, Kawahara D, Kurohashi S. Morphological analysis for unsegmented languages using recurrent neural network language model. 2015 Presented at: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing; 17-21 September 2015; Lisbon, Portugal p. 2292-2297 [doi: [10.18653/v1/D15-1276](https://doi.org/10.18653/v1/D15-1276)]
28. Shibata T, Kawahara D, Kurohashi S. Improving accuracy of Japanese parsing with BERT [in Japanese]. 2019 Presented at: Proceedings of the 25th Annual Conference of the Association for Natural Language Processing; March 12-15, 2019; Nagoya, Japan p. 205-208
29. Kobayashi N, Inui K, Matsumoto Y, Tateishi K, Fukushima T. Collecting evaluative expressions for opinion extraction. 2004 Presented at: Natural Language Processing – IJCNLP 2004; 22-24 March, 2004; Hainan Island, China p. 596-605 [doi: [10.1007/978-3-540-30211-7_63](https://doi.org/10.1007/978-3-540-30211-7_63)]
30. Higashiyama M, Inui K, Matsumoto Y. Learning sentiment of nouns from selectional preferences of verbs and adjectives [in Japanese]. 2008 Presented at: Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing; 17-21 March, 2008; Tokyo, Japan p. 584-587
31. Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 2015;10(3):e0121945 [FREE Full text] [doi: [10.1371/journal.pone.0121945](https://doi.org/10.1371/journal.pone.0121945)] [Medline: [25835001](https://pubmed.ncbi.nlm.nih.gov/25835001/)]
32. Copyright law for information analysis issued by the Japanese Agency for Cultural Affairs [in Japanese]. Agency for Cultural Affairs, Government of Japan. URL: https://www.bunka.go.jp/seisaku/chosakuken/seidokaisetsu/gaiyo/chosakubutsu_jiyu.html [accessed 2023-08-08]
33. Ethical guidelines for life sciences and medical research involving human subjects [in Japanese]. Ministry of Education; Ministry of Health, Labour and Welfare; and Ministry of Economy, Trade and Industry. URL: <https://www.mhlw.go.jp/content/001077424.pdf> [accessed 2023-08-08]
34. Schlesinger M, Grob R, Shaller D. Using patient-reported information to improve clinical practice. *Health Serv Res* 2015;50(Suppl 2):2116-2154 [FREE Full text] [doi: [10.1111/1475-6773.12420](https://doi.org/10.1111/1475-6773.12420)] [Medline: [26573890](https://pubmed.ncbi.nlm.nih.gov/26573890/)]
35. Blödt S, Müller-Nordhorn J, Seifert G, Holmberg C. Trust, medical expertise and humaneness: a qualitative study on people with cancer' satisfaction with medical care. *Health Expect* 2021;24(2):317-326 [FREE Full text] [doi: [10.1111/hex.13171](https://doi.org/10.1111/hex.13171)] [Medline: [33528878](https://pubmed.ncbi.nlm.nih.gov/33528878/)]
36. 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 Inform* 2021;28(1):e100262 [FREE Full text] [doi: [10.1136/bmjhci-2020-100262](https://doi.org/10.1136/bmjhci-2020-100262)] [Medline: [33653690](https://pubmed.ncbi.nlm.nih.gov/33653690/)]
37. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Machine learning and sentiment analysis of unstructured free-text information about patient experience online. *The Lancet* 2012;380(1):S10 [FREE Full text] [doi: [10.1016/S0140-6736\(13\)60366-9](https://doi.org/10.1016/S0140-6736(13)60366-9)]
38. Salgado TM, Liu J, Reed HL, Quinn CS, Syverson JG, Le-Rademacher J, et al. Patient factors associated with discrepancies between patient-reported and clinician-documented peripheral neuropathy in women with breast cancer receiving paclitaxel: a pilot study. *Breast* 2020;51:21-28 [FREE Full text] [doi: [10.1016/j.breast.2020.02.011](https://doi.org/10.1016/j.breast.2020.02.011)] [Medline: [32193049](https://pubmed.ncbi.nlm.nih.gov/32193049/)]
39. Kübler-Ross E, Wessler S, Avioli LV. On death and dying. *JAMA* 1972;221(2):174-179 [FREE Full text] [doi: [10.1001/jama.1972.03200150040010](https://doi.org/10.1001/jama.1972.03200150040010)] [Medline: [5067627](https://pubmed.ncbi.nlm.nih.gov/5067627/)]
40. Ambalavanan AK, Devarakonda MV. Using the contextual language model BERT for multi-criteria classification of scientific articles. *J Biomed Inform* 2020;112:103578 [FREE Full text] [doi: [10.1016/j.jbi.2020.103578](https://doi.org/10.1016/j.jbi.2020.103578)] [Medline: [33059047](https://pubmed.ncbi.nlm.nih.gov/33059047/)]
41. Li X, Yuan W, Peng D, Mei Q, Wang Y. When BERT meets Bilbo: a learning curve analysis of pretrained language model on disease classification. *BMC Med Inform Decis Mak* 2022;21(Suppl 9):377 [FREE Full text] [doi: [10.1186/s12911-022-01829-2](https://doi.org/10.1186/s12911-022-01829-2)] [Medline: [35382811](https://pubmed.ncbi.nlm.nih.gov/35382811/)]
42. Huang K, Altosaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. ArXiv. Preprint posted online on November 29 2020 2020 [FREE Full text]
43. Jenkinson C, Coulter A, Bruster S, Richards N, Chandola T. Patients' experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care. *Qual Saf Health Care* 2002;11(4):335-339 [FREE Full text] [doi: [10.1136/qhc.11.4.335](https://doi.org/10.1136/qhc.11.4.335)] [Medline: [12468693](https://pubmed.ncbi.nlm.nih.gov/12468693/)]

44. Elliott MN, Cohea CW, Lehrman WG, Goldstein EH, Cleary PD, Giordano LA, et al. Accelerating improvement and narrowing gaps: trends in patients' experiences with hospital care reflected in HCAHPS public reporting. *Health Serv Res* 2015;50(6):1850-1867 [[FREE Full text](#)] [doi: [10.1111/1475-6773.12305](https://doi.org/10.1111/1475-6773.12305)] [Medline: [25854292](https://pubmed.ncbi.nlm.nih.gov/25854292/)]
45. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. *BMJ Qual Saf* 2016;25(6):404-413 [[FREE Full text](#)] [doi: [10.1136/bmjqs-2015-004309](https://doi.org/10.1136/bmjqs-2015-004309)] [Medline: [26464518](https://pubmed.ncbi.nlm.nih.gov/26464518/)]

Abbreviations

BERT: bidirectional encoder representation from transformers

ML: machine learning

NLP: natural language processing

PSI: patient satisfaction index

PV: pharmacovigilance

QOL: quality of life

RMSE: root-mean-squared error

SNS: social networking service

Edited by A Mavragani; submitted 01.05.23; peer-reviewed by Z Huang, M Burns; comments to author 16.05.23; revised version received 28.07.23; accepted 09.08.23; published 14.09.23

Please cite as:

Matsuda S, Ohtomo T, Okuyama M, Miyake H, Aoki K

Estimating Patient Satisfaction Through a Language Processing Model: Model Development and Evaluation

JMIR Form Res 2023;7:e48534

URL: <https://formative.jmir.org/2023/1/e48534>

doi: [10.2196/48534](https://doi.org/10.2196/48534)

PMID: [37707946](https://pubmed.ncbi.nlm.nih.gov/37707946/)

©Shinichi Matsuda, Takumi Ohtomo, Masaru Okuyama, Hiraku Miyake, Kotonari Aoki. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.