PEC Innovation 1 (2022) 100017

Contents lists available at ScienceDirect

PEC Innovation

journal homepage: www.elsevier.com/locate/pecinn

# Examining how physician factors influence patient satisfaction during clinical consultations about cancer prognosis and pain

Zhouyang Lou [a], Carolina Vivas-Valencia [b], Cleveland G. Shields [c], Nan Kong [b,*]

[a] School of Industrial Engineering, Purdue University, West Lafayette, IN, USA
[b] Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA
[c] Human Development and Family Studies, Purdue University, West Lafayette, IN, USA

ARTICLE INFO

*Keywords:*
Patient-physician communication
Patient satisfaction
Machine learning
Cancer prognosis
Pain management

ABSTRACT

*Objective:* Patient-physician communication affects cancer patients' satisfaction, health outcomes, and reimbursement for physician services. Our objective is to use machine learning to comprehensively examine the association between patient satisfaction and physician factors in clinical consultations about cancer prognosis and pain.
*Methods:* We used data from audio-recorded, transcribed communications between physicians and standardized patients (SPs). We analyzed the data using logistic regression (LR) and random forests (RF).
*Results:* The LR models suggested that lower patient satisfaction was associated with more in-depth prognosis discussion; and higher patient satisfaction was associated with a greater extent of shared decision making, patient being black, and doctor being young. Conversely, the RF models suggested the opposite association with the same set of variables.
*Conclusion:* Somewhat contradicting results from distinct machine learning models suggested possible confounding factors (hidden variables) in prognosis discussion, shared decision-making, and doctor age, on the modeling of patient satisfaction. Practitioners should not make inferences with one single data-modeling method and enlarge the study cohort to help deal with population heterogeneity.
*Innovation:* Comparing diverse machine learning models (both parametric and non-parametric types) and carefully applying variable selection methods prior to regression modeling, can enrich the examination of physician factors in characterizing patient-physician communication outcomes.

## 1. Introduction

Patient satisfaction is a priority for healthcare systems because patients' HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) evaluations of healthcare providers affect reimbursement for services by the Centers for Medicare and Medicaid (CMS). Patients being satisfied with their care also influences their trust, loyalty and word of mouth on the care, thus impacting patient retention and referrals for providers [1-3]. As a result, care providers have incentives to improve their service delivery to maintain patient satisfaction and remain competitive in the healthcare market. The benefits of improved delivery include better management procedures, prioritizing resource allocations and professional training needs [4,5]. Patients with higher levels of satisfaction are more likely to follow up doctor's appointments or adhere to recommended treatment options [6-8].

In this study, we examined communication behaviors and physician characteristics that we hypothesized would differentiate cancer patients with high satisfaction from the rest and cancer patients with low satisfaction from the rest. Understanding the factors affecting patient satisfaction is the key to improving patient care and satisfaction. There are three categories of factors [9], including non-modifiable predictors, modifiable predictors, and environmental determinants [10]. The non-modifiable predictors include better subjective health [11,12], better functional status [13], or a lower pain level [14], which are associated with higher satisfaction. Patients' race and communication role are also non-modifiable predictors but they are randomized factors in the original study [15]. Since non-modifiable predictors are beyond the control of healthcare providers, they are less helpful in regulating and/or improving patient satisfaction. On the other hand, modifiable predictors suggest avenues for improvement such as patient-centered communication (PCC), time spent waiting to see the physician, and visit duration [14,16-18].

Among these factors, PCC has the strongest impact on patient satisfaction, because of its focus on *eliciting patients' perspectives*, *understanding patients' environments*, *developing a shared understanding of the problem*, and *sharing the decision-making power with patients* [19]. We assessed eliciting patients' perspectives from the communication coding of prognosis and pain

* Corresponding author at: Nan Kong 206 S. Martin Jischke Dr., West Lafayette, IN 47907, USA.
*E-mail address:* nkong@purdue.edu (N. Kong).

http://dx.doi.org/10.1016/j.pecinn.2022.100017
Received 9 August 2021; Received in revised form 7 December 2021; Accepted 3 January 2022

talk, which examined the extent to which physicians explored patients' experiences. Discussions of prognosis and pain were also highly relevant to patients' environment. Finally, we coded for shared decision-making, which examined patients' involvement in treatment discussions and decisions for care. When patients report good communication with their physicians, they are more likely to be satisfied with their care [20]. To physicians, PCC facilitates their comprehension of medical information, promotes better identification of patient needs, perceptions, and expectations [21]. Further, PCC is shown to be an indicator of patients' self-care behavior and their eventual health outcomes, e.g., increased patient adherence to medication and treatment, improved recovery from illness, daily functioning, ability to tolerate pain, and improved patient satisfaction [22-24]. PCC is highly relevant in the care of cancer patients [25].

In this study, we used standardized patients (SPs) to portray stage 4 lung cancer patients with uncontrolled pain. SPs were randomly assigned by race and communication role (activated versus typical) to make unannounced visits to oncologists and primary care physicians who consented to participate in the study [26]. Uncontrolled pain significantly reduces cancer patients' satisfaction with care [27]. Poor communication quality is associated with greater pain intensity and dissatisfaction with care [28]. Palliative care interventions, designed to increase shared decision-making about pain management and prognosis discussions, are associated with improvements in satisfaction [29]. Thus, we hypothesized that SP satisfaction would be positively correlated with pain talk, prognosis talk, and shared decision making.

Several studies investigated the importance of PCC for cancer patient satisfaction with pain management. Hagerty et al. [20] found that when cancer patients report good communication with their physicians, they are more likely to be satisfied with their care. A qualitative study in Finland [30] found that participants emphasized the importance of communication and kindness in addition to a proactive approach to cancer pain assessment. Beck et al. [31] found that factors leading to low satisfaction of cancer patients, included insufficient information about diagnoses, questions not being fully answered, and limited explanation about medications and health conditions. Furthermore, effective pain management relies on productive patient-physician communication [32-35]. In addition, several recent studies suggested that racial disparities in health care are correlated with the differences in patient-physician communication. For example, Penner et al. [36] found that communication within medical interactions significantly influences health disparities in cancer treatment between black and white patients. Studies assessing patient satisfaction via Consumer Assessment of Healthcare Providers and Systems (CAHPS) found that black patients reported lower satisfaction with patient-physician communication [37-39]. Current evidence highlights the importance of training physicians to engage in high-quality communication and minimize the effects of racial discordance by improving patient-centeredness [40].

Previous observational studies of patient-physician communication about cancer prognosis and pain management examined relationships between communication behavior and patient satisfaction via traditional correlation analysis and linear regression [41,42]. However, there is increasing evidence that these statistical techniques have limited power to explore different potential relationships. Additionally, it is commonly seen that non-statisticians use univariate significance (e.g., *t*-test, chi-square test) to select variables for a regression model. However, such approach may reject an important variable when it is confounded by variables that are not controlled [43]. Machine-learning methods, such as random forests, rely less on modeling assumptions (e.g., normally distributed and independent variables), and they can develop models with variables that cannot be entirely characterized by low-degree statistical moments. As a result, machine-learning models may provide greater flexibility to explore nonlinear relationships between communication behavior and patient satisfaction, and subsequently, extend our understanding of these relationships. In addition, we examined the joint impact made by physician communication behavior and their demographic factors on patient satisfaction. Moreover, we showed that the choice of variable selection methods, which is a common procedure prior to regression modeling, could affect the identification of significant independent variables. Overall, this research is intended to demonstrate the use of machine learning to identify significantly influential factors in the context of patient-physician communication and compare diverse machine learning models to reveal the necessary cautions that should be placed in the data-enabled studies.

## 2. Methods

This is a secondary analysis of data from a randomized field trial examining the effects of race and patient activation on opioid prescribing for stage 4 lung cancer patients [26]. We used data from 187 audio-recorded, transcribed visits between physicians and standardized patients (SPs) who portrayed patients with stage 4 lung cancer experiencing uncontrolled pain. In addition, we asked each SP to complete a satisfaction questionnaire after each physician visit in which they rated their satisfaction with the visit.

Because patient satisfaction was skewed toward higher scores, which is common for patient satisfaction measures [44], we transformed patient satisfaction into a 3-level response (i.e., low, medium, and high-level satisfaction). We developed machine learning models for two binary classification tasks to 1) distinguish patients with high-level satisfaction on the patient-physician communication from the rest and 2) distinguish patients with low-level satisfaction from the rest. Our goal is to demonstrate the applicability of machine learning and analyze the differences between machine learning models in examining the influence of physician factors on patient satisfaction from clinical consultation about cancer prognosis and pain.

Machine learning techniques have been widely used to help discover truly complex patterns in data and build accurate prediction models on the outcomes of interest. However, it may suffer from lack of interpretability, its findings on significant variables often raise concerns from the domain experts. We compared two types of machine learning models in a specific healthcare context and selected both logistic regression (a parametric modeling approach) and random forests (a non-parametric modeling approach) for comparison. In addition, variable selection is a commonly used procedure before building regression models. We considered three variable selection methods, namely 1) selection through univariate testing, 2) stepwise variable selection and 3) least absolute shrinkage and selection operator (LASSO).

The actual modeling development is summarized here. We first cleaned raw physician communication behavioral data coded from the transcripts of their clinical consultation with SPs. We then combined the physician behavioral variables together with their demographic variables to form our data set on independent variables. We finalized the data set by appending the patient satisfaction response as the label (i.e., high satisfaction vs. not high; low satisfaction vs. not low) for the supervised binary classification. We also performed data imputation for the missing entries. We next conducted variable selection for the LR modeling. We then developed both models and measured their performance metrics for comparative assessment, including area under the curve (AUC), *p*-value, and adjusted odds ratio. We also visualized the feature influences using a partial dependent plot for the RF models developed.

### *2.1. Data collection*

We analyzed a set of secondary data collected from a Social and Behavioral Influences (SBI) study, which was a randomized field experiment conducted in small metropolitan and rural areas of Indiana, Michigan, and New York [26]. One purpose of the study was to examine physicians" pain assessment and treatment decision-making in advanced lung cancer [45]. Fourteen SPs were trained to assume differentiable roles by their race and activation in the patient-physician communication. For the first role, activated SPs were trained to ask direct and critical questions about their diagnosis, prognosis, and treatment, request information and clarification, and redirect when their concerns were not addressed. In addition, they were provided with a list of questions and instructed to interrupt the physician at least once to ask for clarification. In contrast, those trained as typical

SPs, only asked questions about how to follow through with the treatment, expressed relatively few concerns, appeared satisfied with the information offered, and said they understood even when physician explanations were lacking. Each physician was randomly assigned to see two SPs, both of the same race. The order of activated and typical SPs was random. SP visits were anonymous to the physicians and covertly audio-recorded. Physicians were recruited through direct contact and with the cooperation of a number of large practices and health systems. SPs were trained by experienced SP trainers at each site. Prior to making a visit, SPs performance was rated by naïve raters to be >90%. Each visit was reviewed for adherence to the role and feedback was given to SPs. More information can be found in Shields et al. [45].

***Prognosis related discussion***. We assessed prognosis and treatment choice discussions using the Prognostic Treatment Choices Scale (PTCC), which was developed in a pilot study [46] and recently used in a large randomized intervention trial to improve communication in patients with advanced cancer [47]. These items assess physicians' communication of diagnostic and prognostic information and treatment options with patients with metastatic disease. Sample items are, “Physician asks if the patient wants to know more about the diagnosis” and “Assessing if patients understand their diagnosis.” Coders noted the presence of prognosis items, which we summed for a total prognosis score. Scores of each item were capped at four so that no one item dominated the prognosis total score. The intraclass correlation coefficient (ICC) for prognosis discussions was 0.73.

***Patient-centered pain assessment***. Undergraduate research assistants were trained to code transcripts with the Measure of Physician Pain Assessment (MPPA). The MPPA codes physicians' explorations of patients' worries about pain (e.g., “I understand your medication is not controlling your pain”). This measure was previously developed, piloted and validated for use in outpatient cancer consultations [48]. The ICC for patient centered pain assessment was 0.75.

***Routine pain assessment***. This coded variable assesses routine pain assessment including questions such as pain initiation, location, exacerbating, and alleviating factors [49]. Coders identified pain assessment and pain discussion items that we then summed. Scores on each item were truncated at four so that no one item dominated the total routine pain assessment score. The ICC for routine pain assessment was 0.85.

***Physician implicit association test (IAT) score***. This variable measures a physician's bias toward patients from various racial groups [50-52]. Higher scores mean more racial bias. For example, people can have an implicit preference for white people over black people if they are faster to complete the task when white people + good / black people + bad are paired together compared to when black people + good / white people + bad are paired together.

***Physician word count***. We used the Linguistic Inquiry and Word Count program (LIWC) [53] to calculate the percentage of words physician spoke during a medical visit.

***Physician using “we” statements***. We used the LIWC [53] to calculate the percentage of “we” words. This variable measures the proportion of physicians' we-statements over their all statements in a visit.

***Physician using “I” statements***. We used the LIWC [53] to calculate the percentage of “I” statements physicians made. This variable measures the proportion of physicians' I-statements over their all statements in a visit.

***Shared decision-making***. Shared decision-making was coded using the SDM scale, adapted for chronic conditions from Braddock's Informed Decision Making Scale, which has demonstrated high reliability in several studies in primary care and surgery [54]. The SDM Scale, which identifies nine elements of shared decision-making, has been shown to reliably assess SDM in chronic care (mental health consultations) [55]. The scale includes nine items such as discussion of patient's goals, alternatives for treatment, and exploration of patient's preferences. Each of the nine items were coded as absent (0), partial (1), or complete (2). The ICC for routine pain assessment was 0.53.

***Physician cut-off***. This dichotomous variable assesses whether a physician cuts a patient off by changing the topic in a visit.

***Patient satisfaction***. This response variable was measured with a 36-item 5-point Likert scale that had a Cronbach's alpha = 0.98. The scale consisted of three-items about satisfaction, five-items about empathy, seven-items about physician nonverbal behaviors, and eighteen-items about communication. Each subscale had a Cronbach's alpha $\geq$0.90.

Additionally, we extracted nine variables from patient questionnaires and medical records, including doctor demographics. Physicians ($n$ = 96) were predominantly middle-aged (mean = 52.1, SD = 12.6), white (64%), male (59%), primary care physician (45%), and oncologists (55%). SPs had a range of 8–84 min for each visit. One site was able to enroll only six physicians. Sex was approximately equal for family physicians, but only 32% of oncologists were female.

The summary statistics are reported in Table 1. In this study, we were interested in examining communication and demographic factors that can influence the group of patients with high-level satisfaction from the rest and the group of patients with low-level satisfaction from the rest. We thus regrouped the three levels of patient satisfaction to generate two binary outcomes: 1) low satisfaction (level 1) and not-low satisfaction (levels 2 and 3) and 2) high satisfaction (level 3) and not-high satisfaction (levels 1 and 2). Lastly, we created two interaction terms (i.e., *iat_sp_black4* for the combination of *iat* and *sp_black4*; *iat_sp_active4* for the combination of *iat* and *sp_active4*) since we speculated there could be implicit bias toward either SP type (i.e., being black or being active).

### *2.2. Analysis*

We used the R package ‘MICE’ to impute missing values. Specifically, we implemented predictive mean matching (PMM), a semi-parametric imputation approach [56,57]. It works similarly to regression imputation, except that it fills in a value randomly from a set of candidate donor values for each missing entry. These candidate donor values are drawn from all completing cases whose predictions are closest to the predicted values for the missing entry. We elected to use LR for binary variables and multiple regression for real-numbered variables.

We applied binary LR and RF to model the relationship between patient satisfaction as binary response with quantified communication and demographic factors as exploratory variables or features. With LR, we incorporated three subsets of features from univariate testing, stepwise variable selection, and LASSO variable selection, respectively. With univariate testing, we performed a $t$-test for continuous variables and a chi-square test for

**Table 1**
Summary statistics of sample data.

| Variable | Range | Mean | Std |
|---|---|---|---|
| Prognosis Related Discussion (*prognosis_mean*) | 0 – 8.9 | 1.1 | 0.09 |
| Patient-Centered Pain Assessment (*pcc_explore_pain*) | 3 – 36 | 24.7 | 0.5 |
| Routine Pain Assessment (*pain_mean*) | 0 – 21.5 | 7.8 | 0.3 |
| Physician Implicit Assessment Test (IAT) Score (*iat*) | 0.8 – 1.8 | 0.9 | 0.04 |
| Physician Word Count (*d_percentage_wc*) | 14 – 100 | 68 | 0.008 |
| Physician Using “We” Statements (*d_we*) | 0 – 3.9 | 1.3 | 0.05 |
| Physician Using “I” Statements (*d_i*) | 0 – 5.4 | 2.6 | 0.07 |
| Shared Decision-Making (*sbi_sdm_tot*) | 12 – 45 | 29.5 | 0.6 |
| Physician Age (*age*) | 29 – 80 | 52.1 | 12.6 |
| Visit Minutes (*visit_min*) | 8 – 84 | 33.6 | 15.4 |
| | | Proportion | |
| Physician Cut-Off (*doc_cut_off*) | | 44% | |
| Primary Care Physician (*pcp*) | | 53% | |
| Physician Female (*doc_female*) | | 30% | |
| Physician White (*doc_white*) | | 64% | |
| SP Black (*sp_black*) | | 48% | |
| SP Active (*sp_active*) | | 50% | |
| Patient Satisfaction | level 1 (low) | 22% | |
| | level 2 (med) | 36% | |
| | level 3 (high) | 41% | |

**Table 2**
AUC with LR and RF models.

| | Logistic regression (univariate) | Logistic regression (stepwise) | Logistic regression (LASSO) | Random Forest |
|---|---|---|---|---|
| Low satisfaction | 0.83 | 0.84 | 0.83 | 0.97 |
| High satisfaction | 0.75 | 0.76 | 0.75 | 0.96 |

categorical variables, and then selected exploratory variables with *p* values greater than 0.05. The (forward) stepwise variable selection [58] method started with an LR model with no variables, and iteratively testing if the addition of a variable would decrease the Akaike information criteria (AIC) value, which is a goodness-of-fit measure based on likelihood function and penalized by the number of variables in the model. LASSO variable selection [59] performs both variable selection and regularization to improve model prediction accuracy and interpretability by forcing the summed absolute value of the regression coefficients to be less than a fixed value, resulting from forcing some coefficients to zero to reduce model overfitting. With each subset of selected exploratory variables, we built one LR model for either binary response (i.e., low satisfaction vs. not-low satisfaction and high satisfaction vs. not-high satisfaction). In total, we built six LR models. In addition, we built one RF model with all available data fields as features and either response as above. A RF model randomly generates multiple decision trees with random subsets of features and by bootstrapping the data. Then the RF model can learn from all the decision trees and provide an aggregated prediction. We used R package 'Caret' together with the 'cforest' method to construct the unbiased RF models [60], since continuous and categorical variables together can bias the variable importance measure from RF.

We estimated adjusted odds ratios (ORs) and computed the corresponding 95% confidence intervals (CIs) on each exploratory variable of the LR models. We also computed the area under the receiver operating characteristics (ROC) curve (AUC) as the discrimination measure to compare the constructed LR and RF models. Additionally, we computed variable importance scores for the RF models to compare the features considered more influential by the RF models against the dependent variables tested significantly by the LR models. In the context of RF modeling, variable importance is defined as the increase in the model prediction error when we randomly shuffled the values of the variables. The shuffling breaks the relationship between the features and the response [61]. A variable is considered important if shuffling its values degrades the performance. We further explored their interpretability by using partial dependence plots (PDP) [62]. PDP, proposed by Friedman [63], works well for exploring the interpretability of "black-box" machine learning models. They can show the dependence between the response and desired features while averaged over the values of all other features that are complement. We used R version 4.0.2 to perform all the data analyses.

## 3. Results

Table 2 shows the area under the receiver operating characteristics (ROC) curves (AUC) for the 6 LR models and 2 RF models developed. All the LR models had good discriminatory power for low patient satisfaction (AUC above 0.8), while they only had fair discriminatory power for high patient satisfaction (AUC 0.7–0.8). In contrast, both RF models had almost perfect AUC (0.97 for low satisfaction and 0.96 for high satisfaction). We admit that the perfect AUC is likely attributed to overfitting of the RF models, as is well known in machine learning literature. We emphasize that we used the RF models to offer a perspective potentially as a complement to the LF models in interpreting features associated with patient satisfaction.

Fig. 1 presents the importance scores for all features of the two RF models. All variables were ranked by the descending order of the importance scores from the most to the least important. In contrast to the LR models, the RF models did not provide *p*-values to tell which variables are statistically significant. Instead, variable importance score can only tell what variables are more important in a relative manner. We used an arbitrary threshold of 0.005 to select the relatively more important variables by the RF models for high satisfaction and low satisfaction. With the RF model of low satisfaction, five variables were above the threshold in descending order: patients being black (*sp_black*), depth of discussion about prognosis (*prognosis_mean*), implicit bias toward black people (*iat_sp_black*), doctor word count percentage (*d_percent_wc*), doctor age (*doc_age*), and visit duration (*visit_min*). The variable depth of prognosis

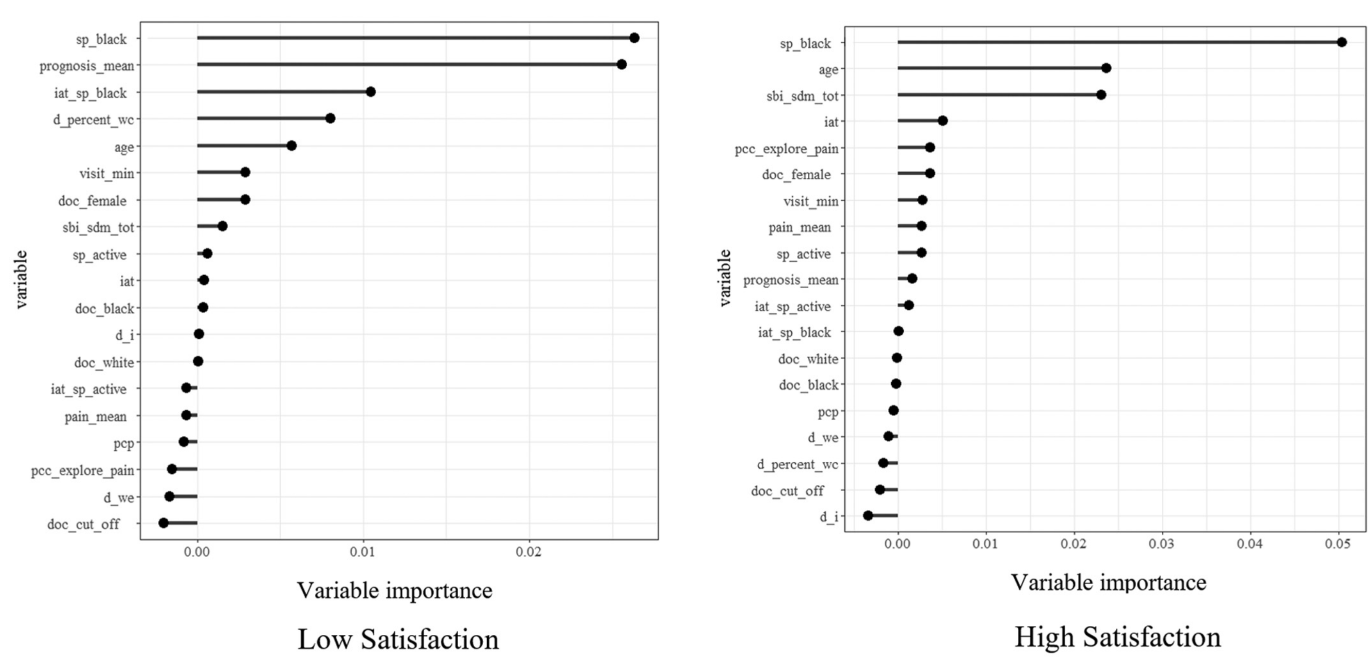


**Fig. 1.** Variable importance.

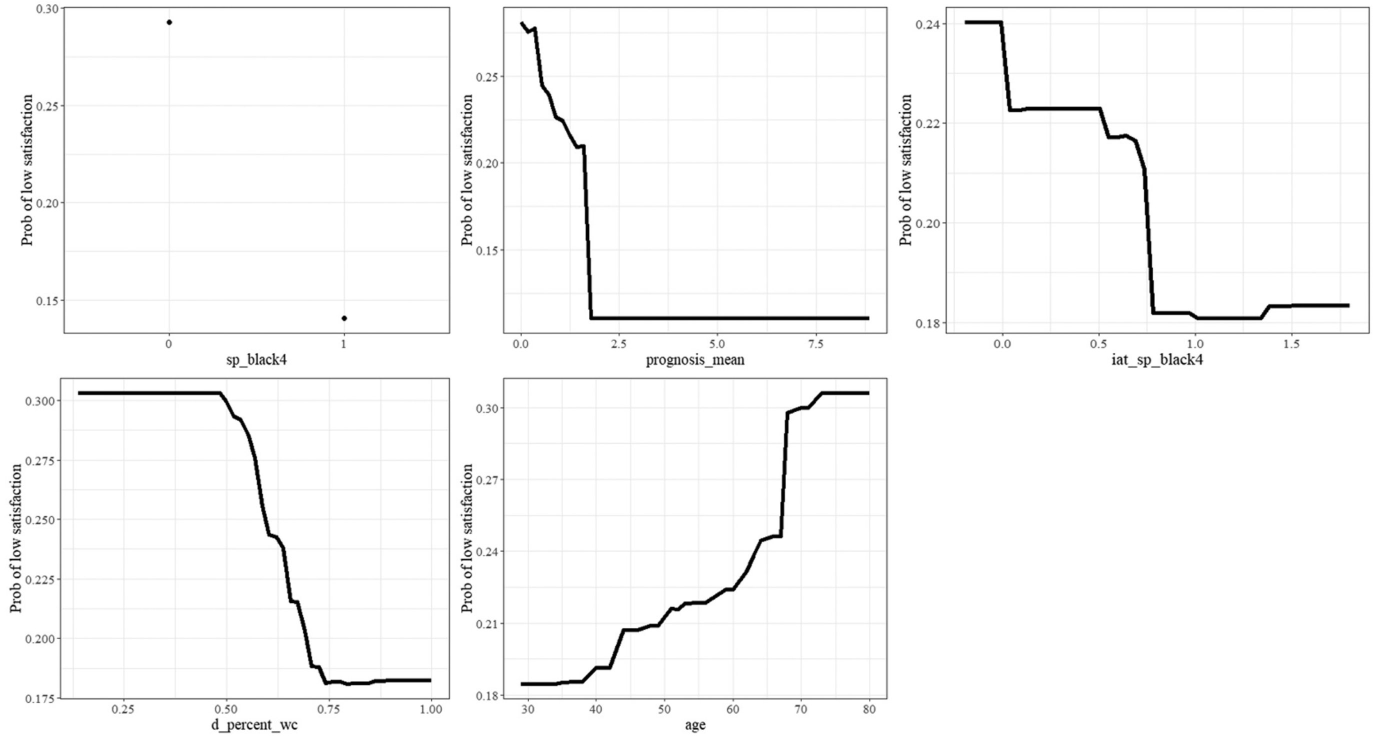


**Fig. 2.** Partial dependence plot (low satisfaction).

discussion was recognized as relatively important and statistically significant by both modeling methods. With the RF model of high satisfaction, four variables exceeded a threshold of 0.005 (descending order): patients being black (*sp_black*), doctor age (*doc_age*), extent of making the decision jointly (*sbi_sdm_tot*), and implicit bias toward patient race (*iat*). Both RF and LR models suggested three factors, patient being black, doctor age, and degree of shared decision-making, would be relatively important or statistically significant.

Figs. 2 and 3 present the partial dependence plots, which show dependence between the response and a set of desired variables (usually one or

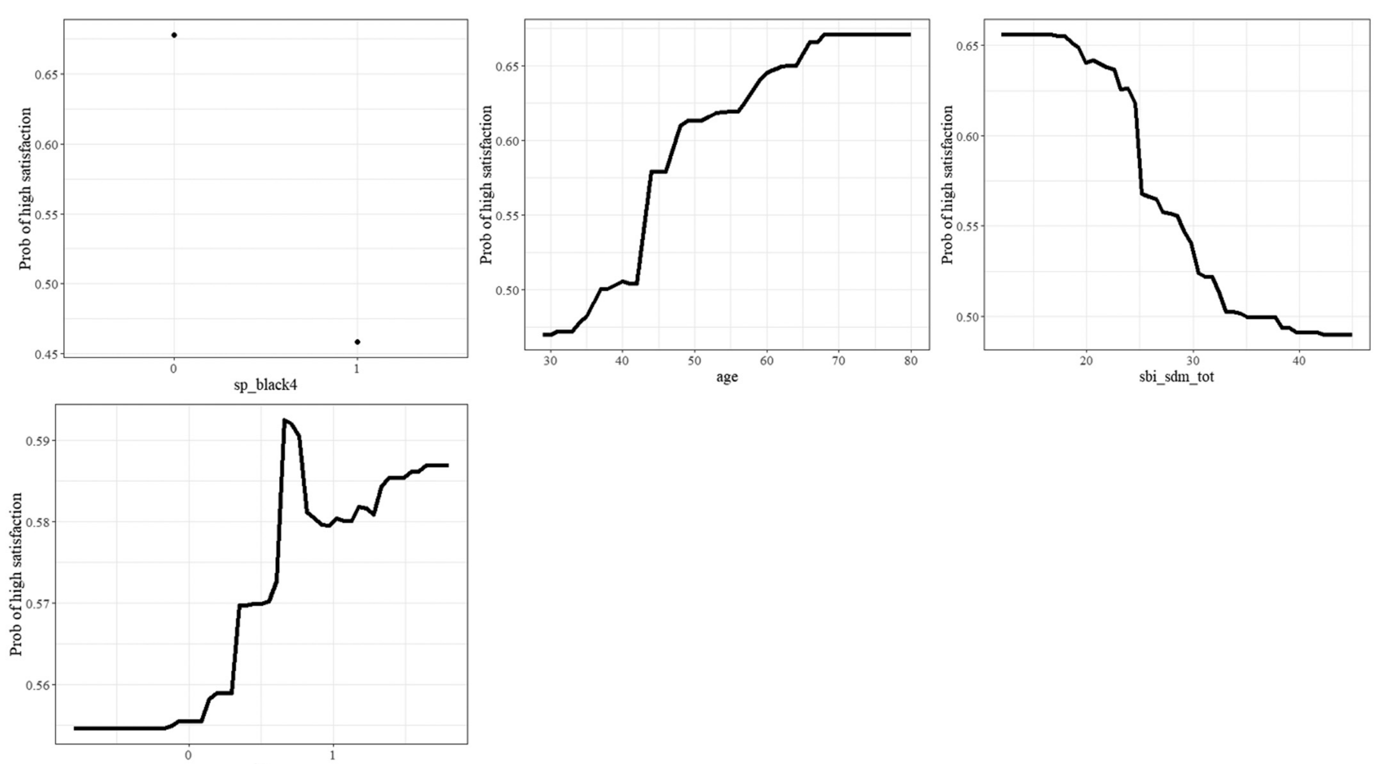


**Fig. 3.** Partial dependence plot (high satisfaction).

two) while marginalizing over the values of all other variables that complement the chosen variables. We showed the top five variables in terms of the RF variable importance score for low satisfaction and the top four variables for high satisfaction, respectively. The x-axis shows the value of each top variable and the y-axis shows the prediction on a patient's probability of becoming quite satisfied or quite unsatisfied. Fig. 2 suggested the following for characterizing low patient satisfaction. One, white patients had a probability 15% higher on being quite unsatisfied compared to black patients. Two, there was a sharp increase on the probability of low satisfaction when the depth of the prognosis discussion dropped below 2.2. Three, a patient's probability of being quite unsatisfied dropped up to 6% when the score of implicit bias toward black patients became greater than 0.6. Four, a patient's probability of getting quite unsatisfied remained at 30% until the doctor's word count percentage during the consultation was above 50%, and dropped to 18% when the doctor word count percentage exceeded 75%. Five, increasing doctor age from 40 to 80 led to a 12% increase in a patient's probability of getting quite unsatisfied, and this probability increased faster for doctors aged 60 years or old. Fig. 3 suggested the following for characterizing high patient satisfaction. One, a white patient increased by 20% her probability of getting quite satisfied compared to a black patient. Two, a patient's probability of getting quite satisfied increased from 45% to 70% when a doctor's age increased from 30 to 80, and the probability increased particularly fast for doctors in their forties. Three, an increased degree of shared decision making led to a decrease on someone's probability of being quite satisfied, i.e., decreasing from 65% to 49%.

Table 3 presents the results from the constructed LR models for the two responses with respect to each of the three variable subsets selected by univariate analysis, stepwise selection, and LASSO, respectively. That is, three LR models for high patient satisfaction and three LR models for low patient satisfaction. A blank cell in the table means that the variable was not selected by the corresponding variable selection method. In applying LR to modeling low patient satisfaction, variables that were selected by all three models (i.e., statistically significant with $p$-value<0.05), included depth of prognosis discussion (*prognosis_mean*, univariate: OR = 2.53, 95% CI 1.43–4.48; stepwise: OR = 2.57, 95% CI 1.45–4.58; LASSO: OR = 2.41, 95% CI 1.35–4.29) and doctors being female (*doc_female*, univariate: OR = 3.68, 95% CI 1.44–9.43; stepwise: OR = 1.28, 95% CI 1.71–10.68; LASSO: OR = 3.57, 95% CI 1.39–9.14). Only in the LR model with stepwise selection, variable patient being black (*sp_black,* OR = 11.71, 95% CI 3.35–40.85) was selected additionally. However, variables *doc_age* and *iat_sp_black*, were not included as they were by the other two variable selection methods. All the three LR models on low satisfaction suggested that more in-depth prognosis discussion and doctors being female were associated with higher odds of low satisfaction. In applying LR to interpreting the high patient satisfaction, variables that were selected by all three models, included extent of making the decision jointly (*sbi_sdm_tot*, univariate: OR = 1.05, 95% CI 1.01–1.09; stepwise: OR = 1.06, 95% CI 1.01–1.1; LASSO: OR = 1.02, 95% CI 0.97–1.08) and patients being black (*sp_black,* univariate: OR = 3.51, 95% CI 1.84–6.69; stepwise: OR = 11.71, 95% CI 3.35–40.85; LASSO: OR = 3.59, 95% CI 1.83–6.92) and doctor age (*doc_age,* univariate: OR = 0.96, 95% CI 0.94–0.99; stepwise: OR = 0.96, 95% CI 0.93–0.99; LASSO: OR = 0.96, 95% CI 0.94–0.99). With stepwise selection, variables visit duration (*visit_min,* OR = 1.02, 95% CI 1–1.04) and implicit bias toward black patients (*iat_sp_black*, OR = 0.3, 95% CI 0.1–0.88) were selected additionally.

All the three LR models on high satisfaction suggested that variable patient being black (*sp_black*) was significantly associated with higher odds of high satisfaction. On the other hand, having a larger extent of making the decision jointly or not (*sbi_sdm_tot*) seemed not affecting the odds of higher satisfaction with ORs close to 1, and so did doctor age (*doc_age*), even though both variables were statistically significant. Additionally, using stepwise selection for the LR modeling, lower implicit bias toward black people (*iat_sp_black*) was associated with lower odds of high satisfaction. All other variables which were not mentioned above were not statistically significant in the corresponding LR model.

## 4. Discussion and conclusion

### 4.1. Discussion

The RF models had better discriminative power than the LR models, as shown by the higher AUCs. However, the RF models offered a different set of influential variables in characterizing low patient satisfaction, compared to the LR models. As for high patient satisfaction, the two modeling methods selected similar influential variables. In addition, we found that a RF model with PDP can augment the quantitative characterization with a fresh perspective. We next interpret our results in the context of clinical communication.

All the models suggested the extent of shared decision-making and doctor age to be influential to differentiating patients with high satisfaction from the rest. The LR models showed that greater extent of shared decision-making and younger doctor age were associated with higher odds of patients being quite satisfied. Conversely, the RF model showed that less shared decision-making and older doctors were associated with a higher probability of patients being quite satisfied. Though the RF model had a better AUC (0.96) compared to the LR models (0.75–0.76), we speculated that the existence of hidden factors in shared decision-making and doctor age might have affected the satisfaction outcomes. One possibility is that doctors may have implemented shared decision-making with patients, but some doctor may only present information to the patients without showing enough empathy or thinking about the patient's perspective. Shared decision-making is a central aspect of PCC [19]. It was positively associated with higher satisfaction according to the RF model. Regardless of how measured, a recent literature review of 39 studies found that shared decision-making was positively associated with patient satisfaction [64]. There is some evidence that patient perceptions of shared decision-making were associated with higher quality of life [65].

**Table 3**
Multivariate logistic regression models with three variable subsets.

| Variables | Low satisfaction OR (95% CI) | | | High satisfaction OR (95%CI) | | |
|---|---|---|---|---|---|---|
| | Logistic regression (univariate) | Logistic regression (stepwise) | Logistic regression (LASSO) | Logistic regression (univariate) | Logistic regression (stepwise) | Logistic regression (LASSO) |
| *prognosis_mean* | 2.53 (1.43–4.48)** | 2.57 (1.45–4.58)** | 2.41 (1.35–4.29)** | – | – | – |
| *pain_mean* | – | – | – | 1.04 (0.97–1.11) | – | 1.04 (0.97–1.11) |
| *sbi_sdm_tot* | – | – | 1.02 (0.97–1.08) | 1.05 (1.01–1.09)* | 1.06 (1.01–1.1)** | 1.05 (1.01–1.1)* |
| *sp_black* | 4.01 (0.7–23.05) | 6.6 (2.7–16.16)*** | 4.01 (0.69–23.23) | 3.51 (1.84–6.69)*** | 11.71 (3.35–40.85)*** | 3.59 (1.86–6.92)** |
| *sp_active* | – | – | – | – | 0.59 (0.3–1.14) | 0.64 (0.34–1.2) |
| *doc_female* | 3.68 (1.44–9.43)** | 4.28 (1.71–10.68)** | 3.57 (1.39–9.14)** | 1.64 (0.87–3.1) | 1.91 (0.98–3.73) | 1.57 (0.83–2.99) |
| *doc_age* | 0.98 (0.94–1.01) | – | 0.98 (0.94–1.02) | 0.96 (0.94–0.99)* | 0.96 (0.93–0.99)* | 0.96 (0.94–0.99)* |
| *visit_min* | – | – | – | – | 1.02 (1–1.04)* | 1.01 (1–1.03) |
| *d_percent_wc* | 18.55 (0.66–524.04) | 21.58 (0.86–541.05) | 16.46 (0.58–469.07) | – | 0.12 (0.01–2.21) | – |
| *iat_sp_black* | 1.69 (0.35–8.24) | – | 1.84 (0.37–9.22) | | 0.3 (0.1–0.88)* | – |
| AIC | 168 | 166 | 366 | 249 | 266 | 228 |

*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; AIC = Akaike Information Criterion.

The doctor age effect may be a result of increases in communication training during medical school. At the same time, younger physicians with less experience may also lead to lower patient satisfaction. On the other hand, older doctors may have plenty of experience but may not be skillful communicators. The literature provides contradictory evidence, patients reported greater levels of satisfaction with older physicians [66] and greater levels of satisfaction with younger physicians [67]. We found that our SPs were less likely to report high satisfaction with older physicians, however, the odds ratios were very close to 1.

Longer visits were found to be associated with greater patient satisfaction [68,69]. However, other studies also found that perceived visit length was not associated with patient satisfaction [70]. In our stepwise LR analysis, visit length was associated with higher satisfaction, but similar to age, the odds ratio was close to 1.

All the models suggested that patient race was influential. The two RF models further suggested that white patients had a higher probability of being quite satisfied or quite unsatisfied. Oppositely, the LR models for high satisfaction showed that black patients were associated with higher odds of being quite satisfied. Furthermore, our results did not show enough evidence that doctor's implicit bias toward patient race was associated with lower patient satisfaction. The conflicting phenomena suggested that hidden racial factors might have affected satisfaction. Given the high percentage of white physicians, it is no surprise that black patients reported lower satisfaction according to the LR model. Given the history of mistreatment and medical mistrust in the black community [71], many patients reported difficulties in trusting and connecting with physicians who might not have their concerns taken seriously [72]. However, according to the RF model, black SPs reported higher levels of satisfaction. However, the RF model also included the interaction of implicit bias and black race. This suggests that the negative effect of race on patient satisfaction may be confined to patients seeing physicians with higher implicit bias.

All the models developed for predicting low patient satisfaction identified the depth of discussion about prognosis to be influential. However, the interpretations from RF model and the LR models are in conflict. The RF model suggested that not enough prognosis discussion might lead to a higher probability of having low patient satisfaction. After the depth of the discussion reaches certain level, the probability of having a low satisfaction level would instead stabilize. Conversely, the LR models suggested that more in-depth prognosis discussion would result in higher odds of low satisfaction. Given a significantly better AUC achieved by the RF model, we would be cautious about trusting the interpretations from LR even though it is commonly used in hypothesis testing in social science. We speculated that some unobservable variables may have contributed to the discrepancy between the two data modeling methods. Oncologists reported resisting the discussions of prognosis because they were afraid of such discussions may disrupt their relationship with patients and families [73]. Patients and family members who reported prognosis discussions with their physicians were slightly more satisfied with their care [74]. In a large study of cancer patient and physician communication, greater levels of prognosis discussion were found to be associated with higher patient reported connection with their physician [75]. However, a recent literature review found no positive associations of prognosis disclosure with the physician-patient relationship [76]. The negative association we found between prognosis and patient satisfaction may be a function of our SPs experiencing prognosis decisions differently than real cancer patients.

The LR models suggested that female doctors would lead to higher odds of low satisfaction. While male and female physicians interact with patients differently, female physicians tend to rate higher on patient-centered communication, but the data on patient satisfaction is mixed. However, the RF model did not deem doctor's gender as a top-five influential feature. Note that all our SPs were male. Physician-patient gender concordance was associated with greater satisfaction [77]. A literature review found that there were no gender differences in overall satisfaction with care, however, in general, patients prefer physicians of their own gender [78]. We speculated that unobserved factors associated with female doctors contributed to low patient satisfaction. In addition, patients being black was shown influential to differentiating low satisfaction by several models.

Our study has several limitations. First, there are missing entries in around one hundred records. We used multiple imputation to address missing data, however, missing data affects the precision of estimates and could result in type II error. Second, our limited SP sample and our programed SPs interactions with physicians may not be representative of the general patient population. Finally, there are limited tools available to interpret black-box machine learning models. The PDP tool for RF models requires making strong assumptions on feature independence, which may not be justifiable in our study. We will address these limitations in our future study, e.g., improving the study protocol design to reduce the missing entries, increasing the sample size of SPs, and utilizing other machine learning models.

### 4.2. Innovation

Machine learning models have emerged as promising methods for modeling observational data and deciphering hidden relationships/patterns in social and behavioral data. They have been widely applied in many scientific areas. Recent literature has suggested transformative potential of machine learning in analyzing data for social science research. Our study showcases such potential for the particular purpose of examining how physician communication behavior factors together with their demographic characteristics jointly influence patient satisfaction on patient-physician communication. Specifically, we believe that our innovation arose from comparing different machine learning models (both regression and random forest) and various variable selection techniques prior to regression modeling. As such, one may acquire additional understanding on what physician factors influence patient outcomes and how.

The use of two different statistical modeling techniques allows for more in-depth analysis of potential predictors of satisfaction. While the two models contradict each other, they suggest the likelihood that unmeasured variables may be at work. The benefit of this information is that PCC researchers, guided by theory, can design future studies to address likely influential variables to include in future models. We believe our analysis demonstrates how machine learning can contribute to understanding influential factors to cancer patients' satisfaction, and we hope that our study will encourage researchers in this area to develop machine learning models when analyzing data which may lead to additional insights.

### 4.3. Practical implication

Our study revealed the complexity of predicting patients' satisfaction with their physicians. Comparing our results to the literature underscored that predictors of satisfaction may be different from study to study. We have suggested that confounding (unmeasured) factors exist that may account for the conflicts in the literature and our own conflicting results. Individuals who examine patient satisfaction such as domain scientists, clinical practitioners, and hospital administrators should not rely on one analytic result obtained from a single model. Even with the results from multiple models, scientists, practitioners, and administrators should be cautious about interpreting factors that have conflicting results from different models. The conflicting results implied that patient satisfaction was not simply associated with patient race, doctor gender, process (e.g., how much shared decision-making? how much prognosis discussion?) or topics in pain management communication. Studies may need to include variables from additional communication channels (such as vocal tone, facial expression, and body language) to capture the full scope of PCC [19].

### 4.4. Conclusion

In this paper, we comprehensively assessed the use of RF modeling with PDP as an additional tool to characterize physician communication behavioral and demographic factors on patient satisfaction from clinical consultation about cancer prognosis and pain. We extracted the communication data from the conversational scripts of the consultation sessions with standard patients. The RF models on patient's low and high satisfaction responses had higher AUC than the corresponding LR models but often resulted in contrary interpretation of certain variables, including the depth of prognosis discussion, the extent of shared decision-making, patient race, and doctor age. For example, low levels of cancer patients' satisfaction were found to be associated with more discussion of prognosis by LR whereas associated with interaction of implicit bias with being a black SP by RF. In the future, we plan to expand the sample size of our cohort study and acquire additional possible confounding factors, e.g., acoustic features and body language among non-verbal sources. As a result, we may gain a more comprehensive understanding of how patient satisfaction can be improved.

## Declaration of Competing Interest

The data used in the study was obtained from the Social and Behavioral Influences Study (SBI) funded by the National Cancer Institute (NCI) grant # 1R01CA155376 to Cleveland G Shields and Ronald E Epstein.

All authors declare that they do not have a conflict of interest.

## References

[1] Kessler DP, Mylod D. Does patient satisfaction affect patient loyalty? Int J Health Care Qual Assur. 2011;24:266–73. https://doi.org/10.1108/09526861111125570.
[2] Noyan F, Şimşek GG. The antecedents of customer loyalty. Procedia Soc Behav Sci. 2014;109:1220–4. https://doi.org/10.1016/J.SBSPRO.2013.12.615.
[3] Ranaweera C, Prabhu J. On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth. J Target Meas Anal Mark. 2003;12:82–90. https://doi.org/10.1057/PALGRAVE.JT.5740100.
[4] Andaleeb S. Service quality perceptions and patient satisfaction: a study of hospitals in a developing country. Soc Sci Med. 2001;52:1359–70. https://doi.org/10.1016/S0277-9536(00)00235-5.
[5] Mpinga E, Chastonay P. Satisfaction of patients: a right to health indicator? Health Policy. 2011;100:144–50. https://doi.org/10.1016/J.HEALTHPOL.2010.11.001.
[6] DuPree E, Anderson R, Nash I. Improving quality in healthcare: start with the patient. Mt Sinai J Med N Y. 2011;78:813–9. https://doi.org/10.1002/MSJ.20297.
[7] Pascoe G. Patient satisfaction in primary health care: a literature review and analysis. Eval Program Plann. 1983;6:185–210. https://doi.org/10.1016/0149-7189(83)90002-2.
[8] Rosenthal G, Shannon S. The use of patient perceptions in the evaluation of health-care delivery systems. Med Care. 1997;35. https://doi.org/10.1097/00005650-199711001-00007.
[9] Shirley E, Sanders J. Patient satisfaction: implications and predictors of success. J Bone Jt Surg. 2013;95. https://doi.org/10.2106/JBJS.L.01048.
[10] Arneill AB, Devlin AS. Perceived quality of care. The influence of the waiting room environment. J Environ Psychol. 2002;22:345–60. https://doi.org/10.1006/JEVP.2002.0274.
[11] Rahmqvist M. Patient satisfaction in relation to age, health status and other background factors: a model for comparisons of care units. International J Qual Health Care. 2001; 13:385–90. https://doi.org/10.1093/INTQHC/13.5.385.
[12] Nguyen Thi P, Briançon S, Empereur F, Guillemin F. Factors determining inpatient satisfaction with care. Soc Sci Med. 2002;54:493–504. https://doi.org/10.1016/S0277-9536(01)00045-4.
[13] Jackson J, Chamberlin J, Kroenke K. Predictors of patient satisfaction. Soc Sci Med. 2001;52:609–20. https://doi.org/10.1016/S0277-9536(00)00164-7.
[14] Säilä T, Mattila E, Kaila M, Aalto P, Kaunonen M. Measuring patient assessments of the quality of outpatient care: a systematic review. J Eval Clin Pract. 2008;14:148–54. https://doi.org/10.1111/J.1365-2753.2007.00824.X.
[15] Hall J, Dornan M. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. Soc Sci Med. 1990;30:811–8. https://doi.org/10.1016/0277-9536(90)90205-7.
[16] Camacho FT, Anderson RT, Safrit A, Jones AS, Hoffmann P. The relationship between patient's perceived waiting time and office-based practice satisfaction. BMC Health Serv Res. 2006;67:409–13. https://doi.org/10.1186/1472-6963-7-31.
[17] Anderson R, Barbara A, Feldman S. What patients want: a content analysis of key qualities that influence patient satisfaction. J Med Pract Manage. 2007;22:255–61.
[18] Young PC, Wasserman RC, McAullife T, Long J, Hagan JF, Heath B. Why families change pediatricians: factors causing dissatisfaction with pediatric care. Am J Dis Child. 1985; 139:683–6. https://doi.org/10.1001/ARCHPEDI.1985.02140090045024.
[19] Epstein RM, Franks P, Fiscella K, Shields CG, Meldrum SC, Kravitz RL, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. Soc Sci Med. 2005;61:1516–28. https://doi.org/10.1016/j.socscimed.2005.02.001.
[20] Hagerty RG, Butow PN, Ellis PM, Dimitry S, Tattersall MHN. Communicating prognosis in cancer care: a systematic review of the literature. Ann Oncol. 2005;16:1005–53. https://doi.org/10.1093/annonc/mdi211.
[21] Roter DL. Communication patterns of primary care physicians. JAMA. 1997;277:350. https://doi.org/10.1001/jama.1997.03540280088045.
[22] Brédart A, Bouleuc C, Dolbeault S. Doctor-patient communication and satisfaction with care in oncology. Curr Opin Oncol. 2005;17:351–4. https://doi.org/10.1097/01.cco.0000167734.26454.30.
[23] Greenfield S, Kaplan SH, Ware JE, Yano EM, Frank HJL. Patients' participation in medical care - effects on blood sugar control and quality of life in diabetes. J Gen Intern Med. 1988;3:448–57. https://doi.org/10.1007/BF02595921.
[24] Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract. 2000;49:796–804.
[25] Epstein RM, Street Jr RL. Patient-centered communication in cancer care: promoting healing and reducing suffering. Bethesda, MD: National Cancer Institute; 2007. NIH Publication No. 07-6225.
[26] Elias CM, Shields CG, Griggs JJ, Fiscella K, Christ SL, Colbert J, et al. The social and behavioral influences (SBI) study: study design and rationale for studying the effects of race and activation on cancer pain management. BMC Cancer. 2017;17. https://doi.org/10.1186/s12885-017-3564-2.
[27] Hannon B, Swami N, Krzyzanowska MK, Leighl N, Rodin G, Le LW, et al. Satisfaction with oncology care among patients with advanced cancer and their caregivers. Qual Life Res. 2013;22:2341–9. https://doi.org/10.1007/s11136-013-0371-3.
[28] Antón A, Montalar J, Carulla J, Jara C, Batista N, Camps C, et al. Pain in clinical oncology: patient satisfaction with management of cancer pain. Eur J Pain. 2012;16:381–9. https://doi.org/10.1002/j.1532-2149.2011.00036.x.
[29] Zimmermann C, Swami N, Krzyzanowska M, Hannon B, Leighl N, Oza A, et al. Early palliative care for patients with advanced cancer: a cluster-randomised controlled trial. The Lancet. 2014;383:1721–30. https://doi.org/10.1016/S0140-6736(13)62416-2.
[30] Simonsen-Rehn N, Sarvimäki A, Sandelin Benkö S. Cancer patients' experiences of care related to pain. Nurs Sci Res Nord Ctries. 2000;57:4–9.
[31] Beck SL, Towsley GL, Berry PH, Lindau K, Field RB, Jensen S. Core aspects of satisfaction with pain management: cancer patients' perspectives. J Pain Symptom Manage. 2010; 39:100–15. https://doi.org/10.1016/J.JPAINSYMMAN.2009.06.009.
[32] Butow P, Hoque E. Using artificial intelligence to analyse and teach communication in healthcare. Breast. 2020;50:49–55. https://doi.org/10.1016/j.breast.2020.01.008.
[33] Matthias MS, Bair MJ, Nyland KA, Huffman MA, Stubbs DL, Damush TM, et al. Self-management support and communication from nurse care managers compared with primary care physicians: a focus group study of patients with chronic musculoskeletal pain. Pain Manag Nurs Off J Am Soc Pain Manag Nurses. 2010;11:26–34. https://doi.org/10.1016/J.PMN.2008.12.003.
[34] Rhiner M. The American Pain Foundation TARGET chronic pain initiative: better patient/clinician communication to improve pain management. J Pain Palliat Care Pharmacother. 2004;18:55–63. https://doi.org/10.1300/J354v18n04_06.
[35] Simon LS. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Mil Med. 2016;181:397–9. https://doi.org/10.7205/MILMED-D-16-00012.
[36] Penner LA, Eggly S, Griggs JJ, Underwood W, Orom H, Albrecht TL. Life-threatening disparities: the treatment of black and white cancer patients. Aust J Soc Issues. 2012; 68. https://doi.org/10.1111/j.1540-4560.2012.01751.x.
[37] Zickmund SL, Burkitt KH, Gao S, Stone RA, Rodriguez KL, Switzer GE, et al. Racial differences in satisfaction with VA health care: a mixed methods pilot study. J Racial Ethn Health Disparities. 2015;2:317–29. https://doi.org/10.1007/s40615-014-0075-6.
[38] Paddison CAM, Elliott MN, Haviland AM, Farley DO, Lyratzopoulos G, Hambarsoomian K, et al. Experiences of care among medicare beneficiaries with ESRD: Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey results. Am J Kidney Dis Off J Natl Kidney Found. 2013;61:440–9. https://doi.org/10.1053/j.ajkd.2012.10.009.
[39] Manfredi C, Kaiser K, Matthews AK, Johnson TP. Are racial differences in patient-physician cancer communication and information explained by background, predisposing, and enabling factors? J Health Commun. 2010;15:272–92. https://doi.org/10.1080/10810731003686598.
[40] Shen MJ, Peterson EB, Costas-Muñiz R, Hernandez MH, Jewell ST, Matsoukas K, et al. The effects of race and racial concordance on patient-physician communication: a systematic review of the literature. J Racial Ethn Health Disparities. 2018;5:117–40. https://doi.org/10.1007/s40615-017-0350-4.
[41] Bishop C. Pattern Recognition and Machine Learning. New York (NY): Springer; 2006. https://doi.org/10.1007/978-3-030-57077-4_11.
[42] Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics. 2018;19:270. https://doi.org/10.1186/s12859-018-2264-5.
[43] Heinze G, Dunkler D. Five myths about variable selection. Transpl Int Off J Eur Soc Organ Transplant. 2017;30:6–10. https://doi.org/10.1111/tri.12895.
[44] Hudak P, Wright J. The characteristics of patient satisfaction measures. Spine. 2000;25: 3167–77. https://doi.org/10.1097/00007632-200012150-00012.
[45] Shields CG, Fuzzell LN, Christ SL, Matthias MS. Patient and provider characteristics associated with communication about opioids: an observational study. Patient Educ Couns. 2019;102:888–94. https://doi.org/10.1016/j.pec.2018.12.005.
[46] Shields CG, Coker CJ, Poulsen SS, Doyle JM, Fiscella K, Epstein RM, et al. Patient-centered communication and prognosis discussions with cancer patients. Patient Educ Couns. 2019;77(3):437–42. https://doi.org/10.1016/j.pec.2009.09.006.
[47] Epstein RM, Duberstein PR, Fenton JJ, et al. Effect of a patient-centered communication intervention on oncologist-patient communication, quality of life, and health care utilization in advanced cancer: the VOICE randomized clinical trial. JAMA Oncol. 2017;3 (1):92–100. https://doi.org/10.1001/jamaoncol.2016.4373.
[48] Shields CG, Finley MA, Elias CM, Coker CJ, Griggs JJ, Fiscella K, et al. Pain assessment: the roles of physician certainty and curiosity. Health Commun. 2013;28(7):740–6. https://doi.org/10.1080/10410236.2012.715380.

[49] Hølen JC, Hjermstad MJ, Loge JH, Fayers PM, Caraceni A, De Conno F, et al. Pain assessment tools: Is the content appropriate for use in palliative care? J Pain Symptom Manage. 2006;32:567–80. https://doi.org/10.1016/j.jpainsymman.2006.05.025.

[50] Cooper LA, Roter DL, Carson KA, Beach MC, Sabin JA, Greenwald AG, et al. The associations of clinicians' implicit attitudes about race with medical visit communication and patient ratings of interpersonal care. Am J Public Health. 2012;102:979–87. https://doi.org/10.2105/AJPH.2011.300558.

[51] Nosek BA, Greenwald AG, Banaji MR. The implicit association test at age 7: a methodological and conceptual review. NY: Psychology Press; 2013.https://doi.org/10.4324/9780203783016-13.

[52] Nosek BA, Bar-Anan Y, Sriram N, Axt J, Greenwald AG. Understanding and using the brief implicit association test: recommended scoring procedures. PLoS One. 2014.;9: e110938. https://doi.org/10.1371/journal.pone.0110938.

[53] Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. https://repositories.lib.utexas.edu/handle/2152/31333; 2015. (accessed October 23, 2021).

[54] Braddock C, Hudak PL, Feldman JJ, Bereknyei S, Frankel RM, Levinson W. "Surgery is certainly one good option": quality and time-efficiency of informed decision-making in surgery. J Bone Joint Surg Am. 2008;90:1830–8. https://doi.org/10.2106/JBJS.G.00840.

[55] Fukui S, Salyers MP, Matthias MS, Collins L, Thompson J, Coffman M, et al. Predictors of shared decision making and level of agreement between consumers and providers in psychiatric care. Community Ment Health J. 2014;50:375–82. https://doi.org/10.1007/s10597-012-9584-0.

[56] Heitjan DF, Little RJA. Multiple imputation for the fatal accident reporting system. Appl Stat. 1991;40:13. https://doi.org/10.2307/2347902.

[57] Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation. Comput Stat Data Anal. 1996;22:425–46. https://doi.org/10.1016/0167-9473(95)00057-7.

[58] Yamashita T, Yamashita K, Kamimura R. A stepwise AIC method for variable selection in linear regression. Commun Stat - Theory Methods. 2007;36:2395–403. https://doi.org/10.1080/03610920701215639.

[59] Tibshirani R. Regression shrinkage and selection via the LASSO. J R Stat Soc Ser B Methodol. 1996;58:267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[60] Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinforma. 2007;81(8):1–21. https://doi.org/10.1186/1471-2105-8-25. (2007).

[61] Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.

[62] Zhao Q, Hastie T. Causal interpretations of black-box models. J Bus Econ Stat. 2021;39: 272–81. https://doi.org/10.1080/07350015.2019.1624293.

[63] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29:1189–232. https://doi.org/10.1214/aos/1013203451.

[64] Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making. 2015;35:114–31. https://doi.org/10.1177/0272989X14551638.

[65] Kashaf MS, McGill E. Does shared decision making in cancer treatment improve quality of life? A systematic literature review. Med Decis Making. 2015;35:1037–48. https://doi.org/10.1177/0272989X15598529.

[66] Chen JG, Zou B, Shuster J. Relationship between patient satisfaction and physician characteristics. J Patient Exp. 2017;4:177–84. https://doi.org/10.1177/2374373517714453.

[67] Bosko T, Wilson K. Assessing the relationship between patient satisfaction and clinical quality in an ambulatory setting. J Health Organ Manag. 2016;30:1063–80. https://doi.org/10.1108/JHOM-11-2015-0181.

[68] Gross DA, Zyzanski SJ, Borawski EA, Cebul RD, Stange KC. Patient satisfaction with time spent with their physician. J Fam Pract. 1998;47:133–7.

[69] Etier BE, Orr SP, Antonetti J, Thomas SB, Theiss SM. Factors impacting Press Ganey patient satisfaction scores in orthopedic surgery spine clinic. Spine J. 2016;16:1285–9. https://doi.org/10.1016/j.spinee.2016.04.007.

[70] Parrish RC, Menendez ME, Mudgal CS, Jupiter JB, Chen NC, Ring D. Patient satisfaction and its relation to perceived visit duration with a hand surgeon. J Hand Surg. 2016;41 (257-262). https://doi.org/10.1016/j.jhsa.2015.11.015. e1–4.

[71] Alsan M, Wanamaker M. Tuskegee and the health of black men. Q J Econ. 2018;133: 407–55. https://doi.org/10.1093/qje/qjx029.

[72] Cuevas AG, O'Brien K, Saha S. African American experiences in healthcare: "I always feel like I'm getting skipped over". Health Psychol Off J Div Health Psychol Am Psychol Assoc. 2016;35:987–95. https://doi.org/10.1037/hea0000368.

[73] Gordon EJ, Daugherty CK. 'Hitting you over the head': oncologists' disclosure of prognosis to advanced cancer patients. Bioethics. 2003;17:142–68. https://doi.org/10.1111/1467-8519.00330.

[74] Heyland DK, Allan DE, Rocker G, Dodek P, Pichora D, Gafni A. Canadian Researchers at the End-of-Life Network (CARENET), discussing prognosis with patients and their families near the end of life: impact on satisfaction with end-of-life care. Open Med Peer-Rev Indep Open-Access J. 2009;3:e101–10.

[75] Fenton JJ, Duberstein PR, Kravitz RL, Xing G, Tancredi DJ, Fiscella K, et al. Impact of prognostic discussions on the patient-physician relationship: prospective cohort study. J Clin Oncol Off J Am Soc Clin Oncol. 2018;36:225–30. https://doi.org/10.1200/JCO.2017.75.6288.

[76] van der Velden NCA, Meijers MC, Han PKJ, van Laarhoven HWM, Smets EMA, Henselmans I. The effect of prognostic communication on patient outcomes in palliative cancer care: a systematic review. Curr Treat Options Oncol. 2020;21:40. https://doi.org/10.1007/s11864-020-00742-y.

[77] Bertakis KD, Franks P, Epstein RM. Patient-centered communication in primary care: physician and patient gender and gender concordance. J Womens Health. 2009;18: 539–45. https://doi.org/10.1089/JWH.2008.0969.

[78] Batbaatar E, Dorjdagva J, Luvsannyam A, Savino MM, Amenta P. Determinants of patient satisfaction: a systematic review. Perspect Public Health. 2017;137:89–101. https://doi.org/10.1177/1757913916634136.