

Language Analysis Predicts Surgical Recommendation Prior to Surgeon Evaluation: An Analysis of 2,954 Pre-Visit Patient Questionnaires

Ryan Halvorson, Kian Niknam, Timothy Samuel Keeley, Brian T Feeley, Alan Zhang¹, Drew Lansdown

¹UCSF Orthopaedic Institute

INTRODUCTION: Patient demand for orthopaedic interventions has been increasing, generating a need for novel strategies to guide resource utilization. Because a large proportion of new clinic patients may be treated conservatively, clinicians must be able to identify patients at risk for requiring intervention in order to anticipate visit duration and predict need for counseling, advanced imaging, or other testing which can delay time to care. Many clinics collect patient questionnaire (PAQ) responses prior to visits to guide discussions. However, these questionnaire responses may be underutilized. Development of advanced language analyses allows for extraction of relevant clinical information from patient free text responses. The purpose of this study was to predict need for surgery for new patients visiting a sports medicine clinic using only pre-visit questionnaire data (i.e. no clinician history, clinical exam, or imaging). We hypothesized that patients requiring surgery could be identified with reasonable accuracy from the questionnaires.

METHODS: *Patient selection.* New patients visiting an orthopaedic sports medicine clinic at a single urban academic center were identified between May 2020 and May 2023. Patients were included retrospectively who completed a pre-visit questionnaire and had a visit with a surgeon. *Study data.* Questionnaire items are shown in Table 1. Patients subsequently undergoing surgery by the same surgeon within 30 days, 90 days, 365 days, and at any time point were identified. Patient age was also obtained from the electronic health record. Aside from patient age, all model features were derived from the PAQ, including pain severity, prior attempt at physical therapy, and response length for several open-ended questions regarding injury, pain quality, and visit expectations (see bolded items in Table 1). Four series of models were generated to classify patients requiring surgery within the four specified time intervals. Models included K nearest neighbors, random forest, logistic regression, gradient boost, and a neural network. All models were trained on a random 80% split of data and tested on a remaining 20% and were evaluated according to accuracy as well as area under the receiver operating characteristic curve (AUC). Post-hoc text analysis was employed to identify specific words or phrases associated with surgical or non-surgical management.

RESULTS: 2,954 patients were identified who completed a pre-visit questionnaire within the study time frame. 249 patients underwent surgery with the same provider within 30 days, 498 within 90 days, 625 within one year, and 1,271 at any timepoint in the future (Figure 1). At each time point, a random forest model generated the highest AUC and the highest accuracy, except for patients undergoing surgery “ever”, who were most accurately predicted by a gradient boost algorithm (Table 2). A random forest model to predict surgical intervention within 30 days had an accuracy of 0.92 and an AUC of 0.67. Within 90 days, a random forest maintained an accuracy of 0.85 with an AUC of 0.63. At 365 days, a random forest was generated with an accuracy of 0.82 and an AUC of 0.62. Finally, patients ever needing surgery were identified with an accuracy of 0.62 and an AUC of 0.66. Patients who had surgery in the first 30 days were more likely to describe symptom chronicity using “days” (6.4% vs 1.8%, $p < 0.001$) and less likely to use “weeks” (3.2% vs. 7.6%, $p = 0.009$), “months” (6.4% vs 15.2%, $p < 0.001$), or “years” (7.5% vs 14.7%, $p < 0.002$). Patients going on to have surgery at any time point were more likely to report having tried at least one prior treatment modality, including anti-inflammatory medication, injections, physical therapy, narcotics, or surgery (21.5% vs 11.3%, $p < 0.001$). Furthermore, patients responding to “what happened” using the word “pop” (11.1% vs 3.9%, $p < 0.001$) or “tear” (6.0% vs 2.3%, $p < 0.001$) were more likely to have surgery within 30 days.

DISCUSSION AND CONCLUSION: In nearly 3,000 patients visiting an orthopaedic sports medicine clinic for the first time, models were able to identify patients requiring operative intervention with reasonable accuracy using only data available prior to the clinic visit (e.g. without clinician history, exam, or imaging). Furthermore, specific words and phrases were identified most predictive of operative intervention for all comers. This suggests language analysis of patient free text data could be employed in combination with machine learning models to triage new patients presenting to clinic. Identifying patients at risk for needing operative intervention would allow clinicians to prioritize time and clinic resources (e.g. schedulers, practice assistants) and dictate urgency for additional imaging or laboratory tests. Models were most reliable for patients requiring acute surgery (e.g. within 30 days). Model accuracy may be improved with the development of more advanced feature extraction from PAQ responses using natural language processing.

Figure 1. Performance of Random Forest to Predict Patients Undergoing Surgery at Various Time Points following Clinic Visit

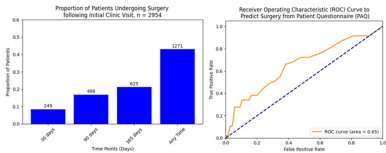


Table 1. Pre-Visit Questionnaire Items	
"What side are we seeing you for?"	
"What part of your body is injured?"	
"When did you start to have pain?"	
"Was there a specific injury?"	
"If so, what happened?"	
"On a scale of 0-10, how severe is your pain?"	
"What treatments have you tried so far?"	
"If you have had a previous surgery or previous surgeries, please list which surgeries you have had below."	
"What makes the pain better?"	
"What makes the pain worse?"	
"What is your current occupation?"	
"What sports/activities do you participate in?"	
"What questions can we answer for you at your appointment?"	

Table 2. Model Performance to Predict Surgery following Initial Clinic Visit							
Model	Within 30 Days		Within 90 Days		Within 365 Days		Ever
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	Accuracy
KNN	0.57	0.92	0.55	0.82	0.56	0.77	0.58
Random Forest	0.67	0.92	0.63	0.85	0.62	0.82	0.62
Logistic Regression	0.56	0.92	0.54	0.85	0.52	0.81	0.59
Gradient Boost	0.61	0.92	0.61	0.84	0.6	0.79	0.63
Neural Net	0.59	0.88	0.62	0.8	0.6	0.75	0.59