

Streamlining the KOOS Activities of Daily Living Subscale Using Machine Learning

Ashim Gupta,^{*†‡§} PhD, MBA, Ajish S.R. Potty,^{*} PhD, Deepak Ganta,^{||} PhD, R. Justin Mistovich,[¶] MD, MBA, Sreeram Penna,[#] MD, Craig Cady,^{**} PhD, and Anish G. Potty,^{*††‡‡} MD

Investigation performed at South Texas Orthopaedic Research Institute, Laredo, Texas, USA

Background: Functional outcome scores provide valuable data, yet they can be burdensome to patients and require significant resources to administer. The Knee injury and Osteoarthritis Outcome Score (KOOS) is a knee-specific patient-reported outcome measure (PROM) and is validated for anterior cruciate ligament (ACL) reconstruction outcomes. The KOOS requires 42 questions in 5 subscales. We utilized a machine learning (ML) algorithm to determine whether the number of questions and the resultant burden to complete the survey can be lowered in a subset (activities of daily living; ADL) of KOOS, yet still provide identical data.

Hypothesis: Fewer questions than the 17 currently provided are actually needed to predict KOOS ADL subscale scores with high accuracy.

Study Design: Cohort study (diagnosis); Level of evidence, 2.

Methods: Pre- and postoperative patient-reported KOOS ADL scores were obtained from the Surgical Outcome System (SOS) data registry for patients who had ACL reconstruction. Categorical Boosting (CatBoost) ML models were built to analyze each question and its value in predicting the patient's actual functional outcome (ie, KOOS ADL score). A streamlined set of minimal essential questions were then identified.

Results: The SOS registry contained 6185 patients who underwent ACL reconstruction. A total of 2525 patients between the age of 16 and 50 years had completed KOOS ADL scores presurgically and 3 months postoperatively. The data set consisted of 51.84% male patients and 48.16% female patients, with a mean age of 29 years. The CatBoost model predicted KOOS ADL scores with high accuracy when only 6 questions were asked ($R^2 = 0.95$), similar to when all 17 questions of the subscale were asked ($R^2 = 0.99$).

Conclusion: ML algorithms successfully identified the essential questions in the KOOS ADL questionnaire. Only 35% (6/17) of KOOS ADL questions (descending stairs, ascending stairs, standing, walking on flat surface, putting on socks/stockings, and getting on/off toilet) are needed to predict KOOS ADL scores with high accuracy after ACL reconstruction. ML can be utilized successfully to streamline the burden of patient data collection. This, in turn, can potentially lead to improved patient reporting, increased compliance, and increased utilization of PROMs while still providing quality data.

Keywords: anterior cruciate ligament; patient-reported outcome measure; Knee injury and Osteoarthritis Outcome Score; function of daily living; activities of daily living; KOOS; ADL; machine learning algorithm; CatBoost

Anterior cruciate ligament (ACL) rupture is a common sports-related injury that, if untreated, can result in continued instability, secondary meniscal tears, and eventual progression toward development of osteoarthritis.^{10,16,17} ACL reconstruction is the standard treatment for ACL ruptures to restore knee biomechanics, allowing for the resumption of prior physical activities and improved quality of life.⁸ To better measure a patient's outcome after ACL reconstruction, surgeons have employed the use of patient-reported outcome measures (PROMs).

The Knee injury and Osteoarthritis Outcome Score (KOOS) is one such knee-specific questionnaire developed to follow patients' functional outcomes after the inciting trauma as well as their recovery postoperatively.¹⁵ Since its initial publication in 1998, the KOOS scale has been extensively used to describe functional outcomes related to knee injuries.¹⁶ Its initial validation study for the English language version of the test was performed using patients undergoing ACL reconstruction. For the past 2 decades, many studies^{4,12,13} have been designed using the KOOS scale as their validated foundation to quantify patient-reported outcomes after ACL reconstruction. KOOS includes 42 questions in 5 distinctly scored subscales, including Pain (9 questions), other Symptoms (7 questions), Activities of Daily

The Orthopaedic Journal of Sports Medicine, 8(3), 2325967120910447
DOI: 10.1177/2325967120910447
© The Author(s) 2020

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

Living (ADL) (17 questions), Function in Sport and Recreation (5 questions), and knee-related Quality of Life (4 questions).¹⁵

While PROMs provide meaningful data, they can be time consuming for patients to complete and resource intensive for practices to administer. The KOOS scale has been validated, but are all of the contained questions necessary to obtain equally meaningful data? Could a streamlined questionnaire provide the same quality data yet minimize the burden of administration and completion? Such improvements in PROMs could lead to increased compliance, better outcome reporting in all practice settings, and improved understanding of patient outcomes.

The goal of this study was to identify a subset of essential questions that can accurately identify outcomes utilizing a machine learning (ML) algorithm. ML is a field of study that uses computer algorithms and statistics to identify complex trends and patterns in the data that may not be easily discernible by humans.³ ML uses data to build empirical/statistical models to describe the behavior of a system. In this study, just the KOOS ADL subscale was used rather than the entire questionnaire for “proof of concept.” We hypothesized that fewer questions are actually needed compared with what is currently provided to predict the KOOS ADL score with high accuracy in patients being treated for ACL ruptures.

METHODS

Data Source

Data for analysis were obtained from the Surgical Outcome System (SOS) global registry, an international patient-reported outcome database maintained by Arthrex. No institutional review board (IRB) approval was required, as SOS global registry is IRB approved and adheres to Health Insurance Portability and Accountability Act (HIPAA) regulations. All SOS global registry users have access to the shared deidentified data.

All patients who underwent ACL reconstruction between 2010 and 2018, had completed the pre- and postsurgery KOOS ADL survey, and had at least 3 months of

follow-up data were included. Patients with missing KOOS ADL survey responses were omitted from the data set. Patient characteristics and procedure-related information such as sex, age at treatment, race and ethnicity, and year of operation were also obtained.

Data Preparation and Model Building

Data processing, analysis, and ML model building were performed using commercially available RStudio. The charts for data analysis were obtained using the “ggplot2” package in RStudio. Categorical Boosting (CatBoost) ML models were built using the “CatBoost” package.⁵ CatBoost is a gradient boosting tool kit, which allows for ordered boosting—a modification of a standard gradient boosting algorithm—that avoids target leakage, and uses a new algorithm for processing categorical features.^{6,14}

The data were randomly split into 2 subsets: a training set with 80% data and a test set with the remaining 20% data. ML models to predict KOOS ADL (with survey responses as predictors) were built using the training data set. Several hyperparameter values were evaluated to identify the “best” model. The choice of the best model was based on minimizing the root mean square error. The performance of the best model was then evaluated on the test data set to gauge its performance on this blind, heldout data set. The hyperparameters used for the final model were as follows: iterations = 500, thread_count = 10, border_count = 32, depth = 5, learning_rate = 0.03, and 12_leaf_reg = 3.5.

One of the outcomes of the CatBoost model is the relative importance of each of the input features (ie, KOOS ADL questions) in explaining the overall prediction. The importance of each feature is determined by calculating the difference in the error with and without that feature in the model. A higher error indicates that the feature is more important, while a lower error indicates less importance. Each of the input features is ranked based on this calculation to get the relative feature importance. A predetermined R^2 value of 0.95 was chosen, as the suggested minimal perceptible clinical improvement in KOOS is 8 to 10 points, and explaining 95% of the variance in KOOS would cover this range of 8 to 10 points adequately.¹⁵

††Address correspondence to Anish G. Potty, MD, South Texas Orthopaedic Research Institute, 2438 Monarch Drive, Suite A 300 #148, Laredo, TX 78045, USA (email: anishpotty@gmail.com).

*South Texas Orthopaedic Research Institute, Laredo, Texas, USA.

†Department of Psychology, Illinois Wesleyan University, Bloomington, Illinois, USA.

‡Future Biologics, Lawrenceville, Georgia, USA.

§BioIntegrate, New York, New York, USA.

¶School of Engineering, Texas A&M International University, Laredo, Texas, USA.

*Department of Orthopaedics, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA.

¶Rothman Orthopaedic Institute, Philadelphia, Pennsylvania, USA.

**Department of Biology, Bradley University, Peoria, Illinois, USA.

††Laredo Sports Medicine Clinic, Laredo, Texas, USA.

Final revision submitted December 6, 2019; accepted December 23, 2019.

One or more of the authors has declared the following potential conflict of interest or source of funding: R.J.M. has received educational support from Arthrex and Depuy, consulting fees from OrthoPediatrics and Philips Healthcare, and hospitality payments from Medical Device Business Services and owns stock/stock options in Right Mechanics. A.G.P. has received educational support from Arthrex/Medinc of Texas and Depuy, grant support from Medtronic, consulting fees from Zimmer Biomet, and hospitality payments from Zimmer Biomet, Stryker, and Smith & Nephew. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval was not sought for the present study.

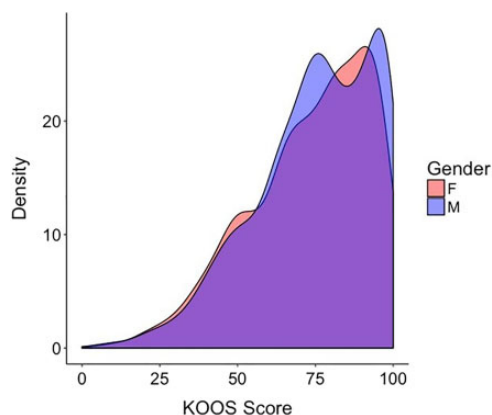


Figure 1. Density plot by count of presurgery KOOS ADL scores for female (F; red) and male (M; blue) in the presurgery data set. KOOS ADL, Knee injury and Osteoarthritis Outcome Score–Activities of Daily Living subscale.

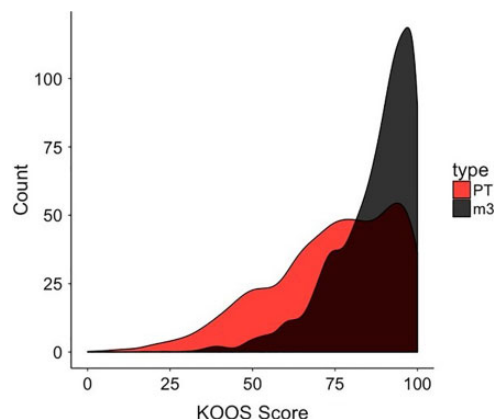


Figure 3. Distribution of KOOS ADL scores presurgery (PT; red) and 3-month postsurgery (m3; gray). KOOS ADL, Knee injury and Osteoarthritis Outcome Score–Activities of Daily Living subscale.

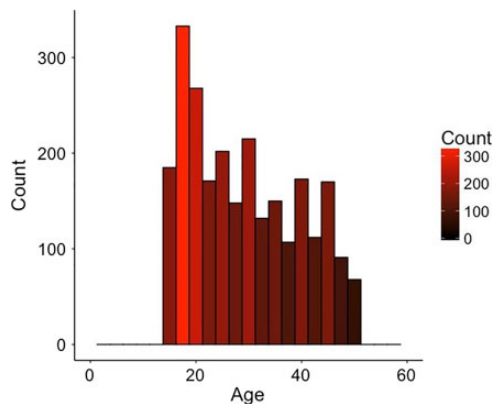


Figure 2. Histogram showing patient age in the presurgery data set.

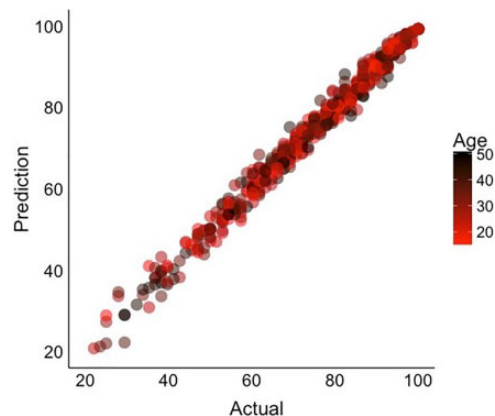


Figure 4. Plot of CatBoost model prediction for test data set using all the 17 questions compared with the actual patient-reported values ($R^2 = 0.99$).

Statistical Analysis

Statistical analysis was performed using RStudio (v 3.4.2). Data for presurgery versus 3 months postsurgery were reported as mean \pm SD for the full KOOS ADL. These values were then compared with the streamlined KOOS ADL values, determining the R^2 value. A Welch 2-sample t test was performed, and $P < .05$ was considered statistically significant.

RESULTS

A total of 6185 patients who underwent ACL reconstruction were identified. Of these patients, 2525 between the ages 16 and 50 years had completed KOOS ADL scores presurgically and 3 months postoperatively. These patients were included in our analysis.

The data set with compliant presurgery responses consisted of 1309 (51.84%) males and 1216 (48.16%) females

(Figure 1). The mean age of the included patients was 29 years (range, 16-50 years) (Figure 2).

The mean \pm SD presurgery and 3-month postsurgery KOOS ADL scores were 73.96 ± 19.31 and 86.66 ± 12.36 , respectively. A Welch 2-sample t test indicated a statistically significant difference ($P < 2.2e-16$) in the means for presurgery and 3-month postsurgery KOOS ADL score (Figure 3). Since the presurgery and 3-month postsurgery KOOS ADL score distributions were statistically different, the 2 distributions were analyzed separately.

The scatter plot of CatBoost model predictions using all 17 questions of the KOOS ADL questionnaire versus the actual scores for the test data set ($n = 505$) predicted KOOS ADL scores with high accuracy ($R^2 = 0.99$) (Figure 4), proving the validity of the initial model.

Interestingly, using the CatBoost model, it was determined that only 6 of the 17 questions (descending stairs, ascending stairs, standing, walking on flat surface, putting

TABLE 1
CatBoost Machine Learning Algorithm Identified
Essential Questions With High KOOS ADL Score
Similar to Full KOOS ADL Questionnaire^a

Question Number	Full KOOS ADL		Streamlined KOOS ADL	
	Questionnaire	R ²	Questionnaire	R ²
A1	Descending stairs	0.99	Descending stairs	0.95
A2	Ascending stairs		Ascending stairs	
A3	Rising from sitting		Standing	
A4	Standing			
A5	Bending to floor/picking up an object			
A6	Walking on flat surface		Walking on flat surface	
A7	Getting in/out of car			
A8	Going shopping			
A9	Putting on socks/stockings		Putting on socks/stockings	
A10	Rising from bed			
A11	Taking off socks/stockings			
A12	Lying in bed (turning over, maintaining knee position)			
A13	Getting in/out of bath			
A14	Sitting			
A15	Getting on/off toilet		Getting on/off toilet	
A16	Heavy domestic duties (moving heavy boxes, scrubbing floors, etc)			
A17	Light domestic duties (cooking, dusting, etc)			

KOOS ADL, Knee injury and Osteoarthritis Outcome Score—Activities of Daily Living subscale.

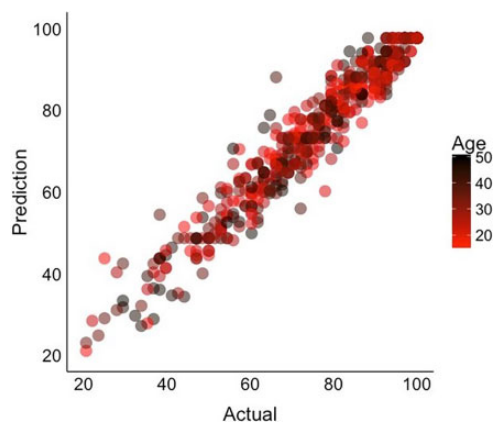


Figure 5. Plot of CatBoost model prediction for test data set using 6 of 17 questions compared with the actual patient-reported values ($R^2 = 0.95$).

on socks/stockings, and getting on/off toilet) (Table 1) were needed to achieve an accuracy in KOOS ADL scores similar to that of the full questionnaire (Figure 5). The specific 6 questions were chosen based on the CatBoost model feature importance (from the training data set) and the ease of quantitatively measuring the response. As shown in Figure 5, the validity of using these 6 questions to predict the KOOS ADL score was confirmed by predicting the KOOS ADL score on the randomly chosen 20% test data set, which was not used in building the model.

DISCUSSION

Functional outcome metrics remain a relatively new yet important public health advancement.¹¹ For much of the history of medicine, clinical success has been defined by the absence of complications or by simple clinical parameters such as range of motion. Despite the validity and usefulness from a research perspective, PROMs pose certain issues, including lengthy questionnaires, redundant questions, and narrow scope, thus limiting their utilization in many practice settings.¹⁷ To circumvent these issues, the Patient-Reported Outcome Measurement Information System (PROMIS) was developed. Implementation of PROMIS led to a significant improvement in the measurement characteristics and a reduction in patient and administrative burden. However, this system was validated only in patients with orthopaedic disorders related to foot and ankle, upper extremities, and spine.¹ Thus, in this study, we focused on using an ML algorithm to identify the important parts of a knee-specific PROM, the KOOS ADL subscale, that has been validated for ACL reconstruction outcomes.

ML adoption is still preliminary in the field of orthopaedics, although in other medical specialties, ML models have been developed and validated to outperform human specialists.^{2,7,9} Nonetheless, the number of publications discussing utilization of ML in orthopaedics since 2000 has increased, indicating its value and potential acceptance in real-world settings.³

The results confirmed the study hypothesis and demonstrated that only 6 questions—descending stairs, ascending stairs, standing, walking on flat surface, putting on socks/stockings, and getting on/off toilet—can reliably predict outcomes with similar accuracy compared with the original 17-question subscale. The use of this abbreviated survey may result in a better patient-reporting experience and compliance while still providing quality data.

Despite encouraging results, this study has several limitations. First, the data were limited to include patients with follow-up data of only 3 months. Future studies will need to include patients with follow-up data of 6 months and 12 months to ensure that the streamlined questionnaire remains equally valid throughout the recovery period. In addition, future studies are needed to evaluate the remaining KOOS subscales, as well as perhaps an even more consolidated generalized full KOOS assessment built from these ML-derived mini-subscales. The successful completion of these studies may lead to the development of a mini-KOOS, with a lower question burden but equal fidelity of data.

CONCLUSION

ML algorithms successfully identified the essential questions in the KOOS ADL questionnaire. Only 35% (6/17) of KOOS ADL questions are needed to predict KOOS ADL scores with high accuracy after ACL reconstruction. Thus, ML can be utilized successfully to streamline the burden of patient data collection. This, in turn, can potentially lead to improved patient reporting, increased compliance, and increased utilization of PROMs, while still providing quality data.

REFERENCES

1. Brodke DJ, Saltzman CL, Brodke DS. PROMIS for orthopaedic outcomes measurement. *J Am Acad Orthop Surg*. 2016;24(11):744-749.
2. Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications. *Science*. 2017;358(6370):1530-1534.
3. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol*. 2018;6:75.
4. Capin JJ, Failla M, Zarzycki R, Dix C, et al. Superior 2-year functional outcomes among young female athletes after ACL reconstruction in 10 return-to-sport training sessions: comparison of ACL-SPORTS randomized controlled trial with Delaware-Oslo and MOON cohorts. *Orthop J Sports Med*. 2019;7(8):2325967119861311.
5. CatBoost Dev Team. CatBoost: R package for CatBoost. R package version 0.8. Yandex Technologies; 2018. <https://tech.yandex.com/catboost/>. Accessed on March 15, 2019.
6. Dorogush AV, Ershov V, Gulin A. *CatBoost: Gradient Boosting With Categorical Features Support*. arXiv preprint arXiv:1810.11363. Cornell University; 2018. <https://arxiv.org/abs/1810.11363>. Accessed March 15, 2019.
7. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.
8. Goebel JC, Bolbos R, Pham M, et al. In vivo high-resolution MRI (7 T) of femoro-tibial cartilage changes in the rat anterior cruciate ligament transection model of osteoarthritis: a cross-sectional study. *Rheumatology*. 2010;49(9):1654-1664.
9. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
10. Gupta A, Sharif K, Walters M, et al. Surgical retrieval, isolation and in vitro expansion of human anterior cruciate ligament-derived cells for tissue engineering applications. *J Vis Exp*. 2014;86:51597.
11. Hamilton DF, Lane JV, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. *BMJ Open*. 2013;3(4):e002525.
12. Ingelsrud LH, Terwee CB, Terluin B, et al. Meaningful change scores in the Knee injury and Osteoarthritis Outcome Score in patients undergoing anterior cruciate ligament reconstruction. *Am J Sports Med*. 2018;46(5):1120-1128.
13. MARS Group, Wright RW, Huston LJ, et al. Predictors of patient-reported outcomes at 2 years after revision anterior cruciate ligament reconstruction. *Am J Sports Med*. 2019;47(10):2394-2401.
14. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Processing Syst*. 2018;31:6639-6649.
15. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1(1):64.
16. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and Osteoarthritis Outcome Score (KOOS)— development of a self-administered outcome measure. *J Orthop Sports Phys Ther*. 1998;28(2):88-96.
17. Zantop T, Petersen W, Sekiya JK, Musahl V, Fu FH. Anterior cruciate ligament anatomy and function relating to anatomical reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2006;14(10):982-992.