Machine Learning Model Developed to Aid in Patient Selection for Outpatient Total Hip Arthroplasty

Columbia | Vagelos College of Physicians & Surgeons
Program for Education in Global and Population Health

Jessica Ding, BA[1]; Cesar D. Lopez, BS[1]; Thomas R. Hickernell, MD[1]
(International Collaboration Exchange Mentor: Anette Wu, MD MPH PhD[1])
[1] Columbia University Vagelos College of Physicians and Surgeons

Research Question: Are machine learning models able to assist with preoperative patient selection for outpatient total hip arthroplasty by using patient comorbidities and demographic factors?

BACKGROUND

Total hip arthroplasty (THA), or total hip replacement, is a common elective surgical procedure that greatly restores quality of life. THA demand is increasing, but it has high costs mainly driven by hospital length of stay (LOS).[1-3] More THA procedures are performed on an outpatient basis, in which the patient has same-day discharge (LOS = 0 days).[4] Outpatient THA is as safe as inpatient THA for appropriately selected patients.[5,6] Selecting patients for outpatient THA by length of stay may help maintain cost effectiveness and good outcomes.

Machine learning (ML) is a form of artificial intelligence (AI) that learns from large datasets to build complex statistical models, which can be used to identify patterns and predict outcomes.[7,8] In orthopedic surgery, ML has been studied for classifying patients and optimizing for surgical procedures,[20] outcome predictions,[21, 22] and patient specific payment models.[23]

The purpose of this study is to develop ML models that may aid in patient selection for outpatient THA by predicting which patients, based on medical comorbidities and demographic information, may have long LOS or same-day discharge.

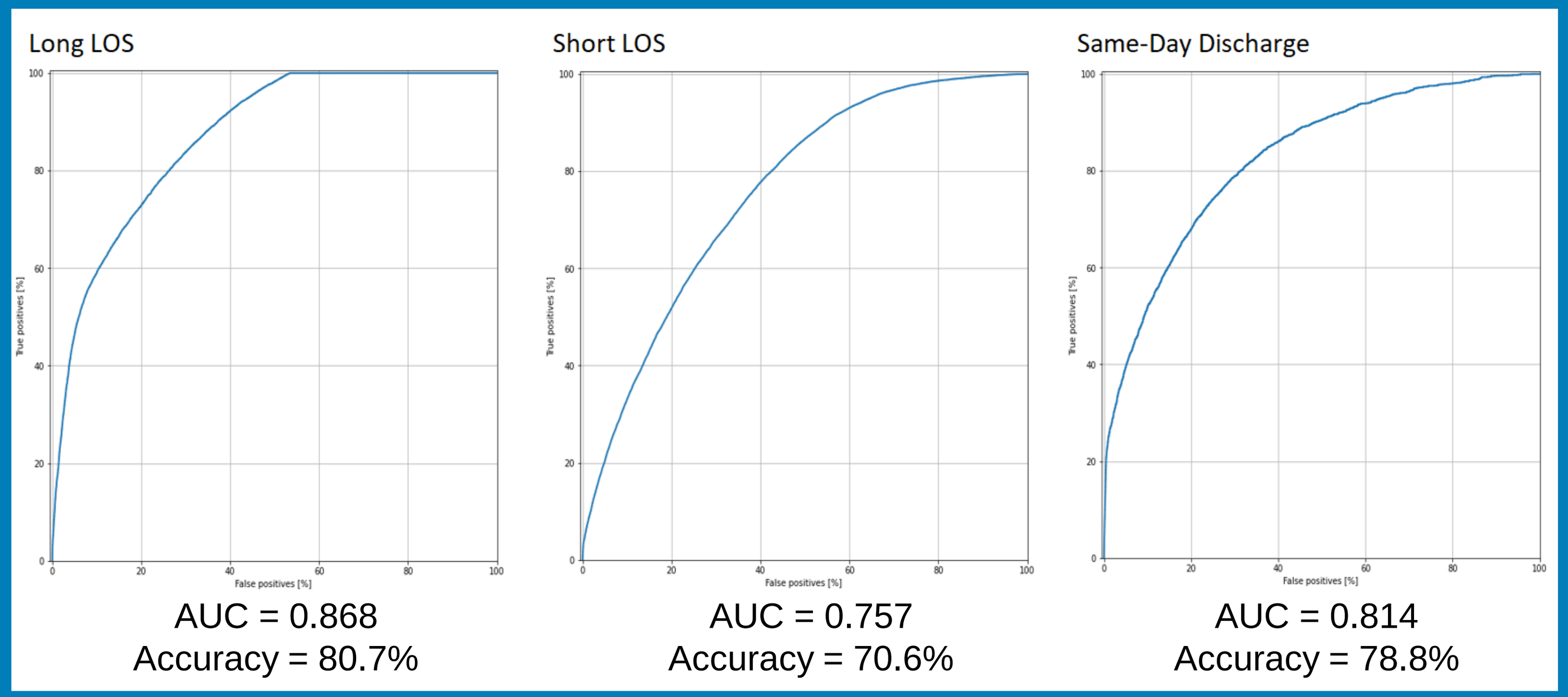
DESCRIPTION OF ORGANIZATION

The **Department of Orthopedic Surgery at New York-Presbyterian Columbia University Irving Medical Center** was formerly the New York Orthopedic Hospital, which was founded for "for the purpose of furnishing treatment to the poor, with special reference to diseases and deformities of the bones and joints requiring surgical and mechanical treatment, and for giving instruction in the same." Today, the Department of Orthopedic Surgery's missions continue through focus on patient care, research, and education.[13]

The **International Collaboration and Exchange Program** was started at Columbia University to provide opportunities for students to network and collaborate with medical students from other countries around the world. Although COVID-19 disrupted travel plans, students planned on participating in summer basic science research rotations in partner countries while experiencing the culture of the partner countries.[14]

RESULTS

Figure 1: Receiver operating characteristic (ROC) curves for machine learning models.



An AUC value of 1.0 is considered perfect discriminative ability with completely correct predictions, while 0.90 to 0.99 is considered excellent, 0.80 to 0.89 is considered good, 0.70 to 0.79 is considered fair, and 0.50 to 0.69 is considered poor.[16]

Figure 2: Odds ratios for factors associated with long LOS (left) and same-day discharge (right).

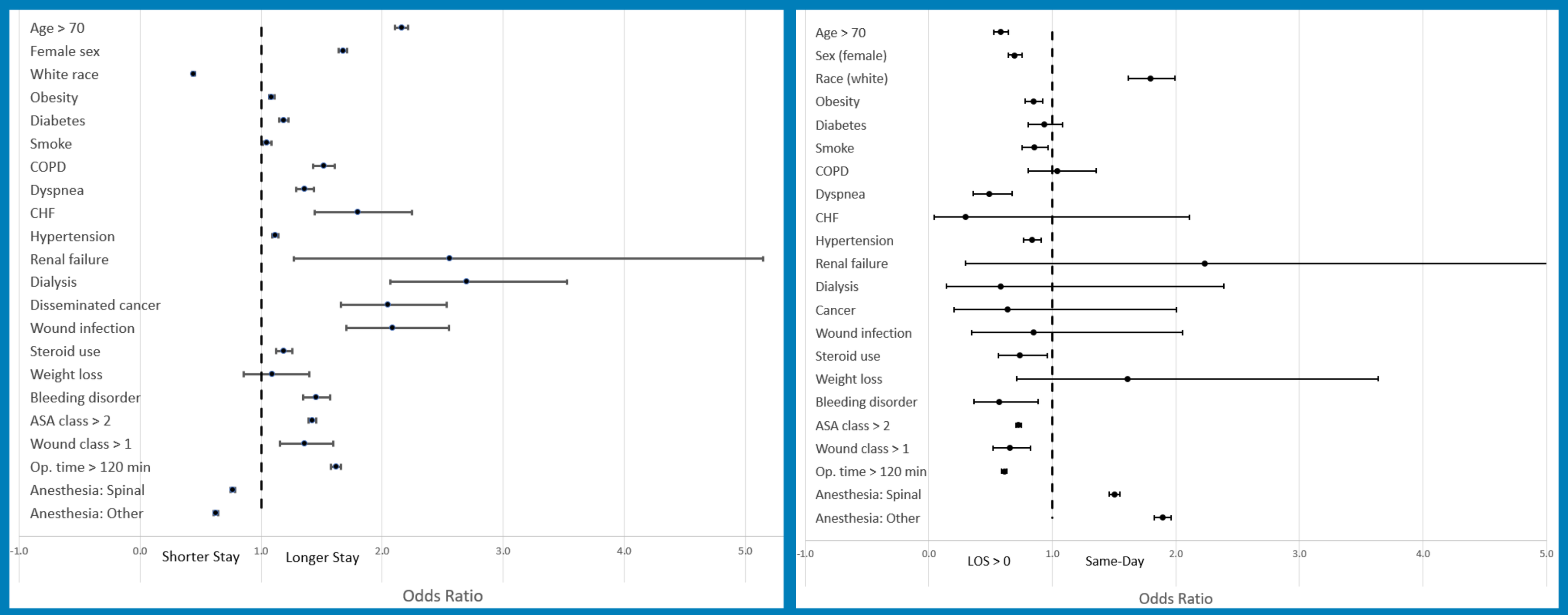


Table 1: Outcomes of cases when categorized by long LOS

Outcomes	Non-Long LOS (n = 84,770)	Long LOS (n = 68,283)	P value	All THA (n= 153,053)
Inpatient (%)	98.5%	99.2%	**< 0.001**	98.8%
Average operative time (min)	89.5	95	**< 0.001**	92.0
Prolonged operative time (%)	14.6%	20.2%	**< 0.001**	17.1%
Average LOS (days)	1.6	3.8	**< 0.001**	2.6
Any complication (%)	3.6%	15.1%	**< 0.001**	8.7%
Non-home discharge (%)	5.8%	37.0%	**< 0.001**	19.8%
Readmission (%)	3.4%	5.0%	**< 0.001**	4.2%
Reoperation (%)	1.7%	2.9%	**< 0.001**	2.3%

Table 2: Outcomes of cases when categorized by same-day discharge

Outcomes	Non-Ambulatory (n =150,432)	Ambulatory (n = 2,621)	P value	All THA (n= 153,053)
Inpatient (%)	99.1%	80.7%	**< 0.001**	98.8%
Average operative time (min)	92.1	82.4	**< 0.001**	92.0
Prolonged operative time (%)	17.3%	9.4%	**< 0.001**	17.1%
Average LOS (days)	2.6	0	**< 0.001**	2.6
Any complication (%)	8.8%	2.6%	**< 0.001**	8.7%
Non-home discharge (%)	20.0%	5.1%	**< 0.001**	19.8%
Readmission (%)	4.2%	3.2%	**0.014**	4.2%
Reoperation (%)	2.3%	2.1%	0.590	2.3%

METHODS

- This retrospective cohort study uses deidentified data from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) registry during 2010-2018.[15]
- Same-day discharge or ambulatory/outpatient THA was defined as LOS = 0 days. Short LOS patients had LOS ≤ 1 day, while long LOS patients had LOS ≥ 3 days.
- Multivariate linear regression and multivariate logistic regression analyses identified factors associated with the outcomes. All statistical analyses were conducted with Stata 16.1 (Stata Corp. – College Station, Texas, USA). Statistical significance was defined as $P < 0.05$.
- Artificial neural network (ANN) models were developed with TensorFlow Python open-source coding platform (Google Brain, Alphabet Inc. – Mountain View, California, USA). The models were trained by randomly sorting the cases into a training set (80%) and a testing set (20%).
- The model's discriminative ability (sensitivity/specificity) was evaluated with the area under the receiver operating characteristic curve (AUC). The overall model accuracy was calculated by dividing the number of correct predictions by the total sample size.

The ML model for predicting long LOS had an AUC of 0.868 and an accuracy of 80.7%. The short LOS ML model had an AUC of 0.757 and an accuracy of 70.6%. The same-day discharge ML model had an AUC of 0.814 and an accuracy of 78.8% (*Figure 1*). Age over 70, female sex, ASA Class of 3 or above, wound class of 2 or above, general anesthesia, prolonged operative time, and other comorbidities were found to increase the odds of a longer stay (*Figure 2*).

DISCUSSION

The ML models predicting long LOS and same-day discharge were considered good predictive and discriminative ability (AUC > 0.80). The model for short LOS had fair discriminative ability (AUC > 0.70). Greater numbers of comorbidities are associated with longer LOS. Non-White race was the main demographic factor found to increase odds of long LOS. These racial disparities likely reflect existing systemic inequalities in the data.[17] Predictive models may inadvertently perpetuate disparities; patients designated as higher risk may have more difficulty in being referred for THA and outpatient THA.
Limitations include lack of external validity, lack of adjustments for surgery chronology between 2010-2018, and database coding errors.

REFERENCES

1. Sloan, M., A. Premkumar, and N.P. Sheth, *Projected Volume of Primary Total Joint Arthroplasty in the U.S., 2014 to 2030.* J Bone Joint Surg Am, 2018. **100**(17): p. 1455-1460.
2. McDermott, K.W., W.J. Freeman, and A. Elixhauser, *Overview of Operating Room Procedures During Inpatient Stays in U.S. Hospitals, 2014: Statistical Brief #233*, in *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. 2006, Agency for Healthcare Research and Quality (US): Rockville (MD).
3. Meyers, S.J., et al., *Inpatient cost of primary total joint arthroplasty.* J Arthroplasty, 1996. **11**(3): p. 281-5.
4. Bert, J.M., J. Hooper, and S. Moen, Outpatient Total Joint Arthroplasty. Curr Rev Musculoskelet Med, 2017. 10(4): p. 567-574.
5. Richards, M., et al., An Evaluation of the Safety and Effectiveness of Total Hip Arthroplasty as an Outpatient Procedure: A Matched-Cohort Analysis. J Arthroplasty, 2018. 33(10): p. 3206-3210.
6. Arshi, A., et al., Outpatient Total Hip Arthroplasty in the United States: A Population-based Comparative Analysis of Complication Rates. J Am Acad Orthop Surg, 2019. 27(2): p. 61-67.
7. Deo, R.C., Machine Learning in Medicine. Circulation, 2015. 132(20): p. 1920-30.
8. Awan, S.E., et al., Machine learning in heart failure: ready for prime time. Curr Opin Cardiol, 2018. 33(2): p. 190-195.
9. Lin, C.C., et al., Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture. Injury, 2010. 41(8): p. 869-73.
10. Fontana, M.A., et al., Can Machine Learning Algorithms Predict Which Patients Will Achieve Minimally Clinically Important Differences From Total Joint Arthroplasty? Clin Orthop Relat Res, 2019. 477(6): p. 1267-1279.
11. Harris, A.H.S., et al., Can Machine Learning Methods Produce Accurate and Easy-to-use Prediction Models of 30-day Complications and Mortality After Knee or Hip Arthroplasty? Clin Orthop Relat Res, 2019. 477(2): p. 452-460.
12. Ramkumar, P.N., et al., Deep Learning Preoperatively Predicts Value Metrics for Primary Total Knee Arthroplasty: Development and Validation of an Artificial Neural Network Model. J Arthroplasty, 2019. 34(10): p. 2220-2227.e1.
13. https://www.columbiaortho.org/about-us/our-history
14. https://www.internationalcollaborationexchange.org/
15. ACS-NSQIP, User Guide for the 2018 ACS NSQIP Participant Use Data File (PUF). 2019.
16. Hanley, J.A. and B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 1982. 143(1): p. 29-36.
17. Sullivan, R., et al., Bundled payments in total joint arthroplasty and spine surgery. Curr Rev Musculoskelet Med, 2017. 10(2): p. 218-223.

Contact info: jd3628@cumc.columbia.edu

Funded by: Columbia University Vagelos College of Physicians and Surgeons | Program for Education in Global and Population Health