



Universiteit
Leiden
The Netherlands

Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty

Karhade, A.V.; Schwab, J.H.; Bedair, H.S.

Citation

Karhade, A. V., Schwab, J. H., & Bedair, H. S. (2019). Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. *Journal Of Arthroplasty*, 34(10), 2272-2277.e1.
doi:10.1016/j.arth.2019.06.013

Version: Publisher's Version
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/4301187>

Note: To cite this publication please use the final published version (if applicable).



Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org



Artificial Intelligence and Machine Learning

Development of Machine Learning Algorithms for Prediction of Sustained Postoperative Opioid Prescriptions After Total Hip Arthroplasty



Aditya V. Karhade, BE, Joseph H. Schwab, MD, MS, Hany S. Bedair, MD *

Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA

ARTICLE INFO

Article history:

Received 16 April 2019

Received in revised form

17 May 2019

Accepted 5 June 2019

Available online 13 June 2019

Keywords:

arthroplasty

machine learning

opioid use

orthopedic surgery

prediction

total hip arthroplasty

ABSTRACT

Background: Postoperative recovery after total hip arthroplasty (THA) can lead to the development of prolonged opioid use but there are few tools for predicting this adverse outcome. The purpose of this study is to develop machine learning algorithms for preoperative prediction of prolonged opioid prescriptions after THA.

Methods: A retrospective review of electronic health records was conducted at 2 academic medical centers and 3 community hospitals to identify adult patients who underwent THA for osteoarthritis between January 1, 2000 and August 1, 2018. Prolonged postoperative opioid prescriptions were defined as continuous opioid prescriptions after surgery to at least 90 days after surgery. Five machine learning algorithms were developed to predict this outcome and were assessed by discrimination, calibration, and decision curve analysis.

Results: Overall, 5507 patients underwent THA, of which 345 (6.3%) had prolonged postoperative opioid prescriptions. The factors determined for prediction of prolonged postoperative opioid prescriptions were age, duration of opioid exposure, preoperative hemoglobin, and preoperative medications (anti-depressants, benzodiazepines, nonsteroidal anti-inflammatory drugs, and beta-2-agonists). The elastic-net penalized logistic regression model achieved the best performance across discrimination (c-statistic = 0.77), calibration, and decision curve analysis. This model was incorporated into a digital application able to provide both predictions and explanations (available at <https://sorg-apps.shinyapps.io/thaopioid/>).

Conclusion: If externally validated in independent populations, the algorithms developed in this study could improve preoperative screening and support for THA patients at high risk for prolonged postoperative opioid prescriptions. Early identification and intervention in high-risk cases may mitigate the long-term adverse consequence of opioid dependence.

Level of Evidence: III.

© 2019 Published by Elsevier Inc.

Disclosures: The authors have nothing to disclose.

Funding: The authors report no funding disclosures for this study.

Ethics Statement: This study was approved by our institutional review board.

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.arth.2019.06.013>.

* Reprint requests: Hany S. Bedair, MD, Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, MA 02114.

<https://doi.org/10.1016/j.arth.2019.06.013>
0883-5403/© 2019 Published by Elsevier Inc.

Americans make up 5% of the global population but consume 80% of the world's opioid supply [1]. From 1999 to 2017, annual deaths related to prescription opioids increased 5-fold with a total of 218,000 deaths over this time period [2]. Opioids are commonly prescribed for the acute recovery from procedures such as total hip arthroplasty (THA) but the postoperative period has been identified as a high-risk setting for the development of long-term opioid dependence [3–6].

Previous studies have identified risk factors for prolonged opioid use after THA, but there are currently no preoperative algorithms to predict this adverse outcome after THA [3,7,8]. Machine learning is a subfield of statistics and computer science with increasing popularity for predictive analytics in orthopedics [9–12]. Previous

studies have used this methodology to predict postoperative opioid use in orthopedic spine surgery but there are no studies that have applied this methodology for opioid prediction in arthroplasty [13].

As such, the purpose of this study is to develop and evaluate machine learning algorithms for preoperative prediction of prolonged opioid prescriptions after THA. An additional aim of this study is to overcome the 2 main limitations of machine learning models: (1) lack of accessibility and (2) lack of interpretability, by deploying the final machine learning algorithm as a digital application and by providing both predictions and explanations.

Materials and Methods

Guidelines

The Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis guidelines and the Journal of Medical Internet Research Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research were followed [14,15].

Data Source

Our institutional review board approved retrospective review of electronic health records for patients undergoing THA at 2 academic medical centers and 3 community hospitals. Inclusion criteria for the study were (1) adult patients, 18 years of age or older; (2) procedure for THA between January 1, 2000 and August 1, 2018; and (3) operative diagnosis of osteoarthritis. Exclusion criteria for the study were revision THA or primary THA for management of fractures, avascular necrosis, rheumatoid arthritis, tumor, infection, or any reason other than osteoarthritis.

Outcome

The primary outcome of interest was prolonged opioid prescription defined based on prior studies as continuous opioid prescriptions after surgery to at least 90 days after surgery [5,16]. Surveillance was conducted through electronic health records and pharmacy records and the list of candidate medications is available for review in [Supplementary Table 1](#).

Variables

Variables included in the analysis were age, gender, race, ethnicity, marital status, veteran status, disposition (inpatient, outpatient), laboratory characteristics [hemoglobin (g/dL) white blood cell count ($\times 10^3/\mu\text{L}$), platelet count ($\times 10^3/\mu\text{L}$), creatinine level (mg/dL)], insurance status, neighborhood characteristics based on the United States Censuses Bureau American Community Survey zip code data [median household income (\$), median age (years), high school graduation and General Equivalency Diploma rate (%), unemployment rate (%), population density (per square mile)], preoperative medications, and comorbidities. Medications were assessed in the year preceding THA and the full list of medications is available for review in [Supplementary Table 1](#). Preoperative duration of opioid use was included as a variable on the basis of prior studies and categorized as continuous opioid use exceeding 180 days, opioid use for less than 180 days, and no opioid use [17,18].

Missing Data

Rates of missing data were race (163, 3.0%), marital status (216, 3.9%), veteran status (126, 2.3%), median household income

(87, 1.6%), high school attainment (67, 1.2%), unemployment rate (68, 1.2%), population density (113, 2.1%), white blood cell count (799, 14.5%), hemoglobin (802, 14.6%), platelet (808, 14.7%), and creatinine (954, 17.3%). The missForest multiple imputation method was used to impute variables with less than 30% missing data [19].

Table 1

Baseline Characteristics of Study Population, n = 5507.

Variable	n (%) Median (IQR)
Age	66.0 (57.0–74.0)
Female gender	2725 (49.5)
Race	
Non-White	294 (5.5)
White	5050 (94.5)
Ethnicity	
Hispanic	40 (0.7)
Non-Hispanic	5304 (99.3)
Marital status	
Married	3289 (62.2)
Not married	2002 (37.8)
Veteran	808 (15.0)
Disposition	
Inpatient	5409 (98.2)
Outpatient	98 (1.8)
Laboratory characteristics	
Hemoglobin	13.6 (12.6–14.6)
White blood cell	7.00 (5.89–8.30)
Platelet	257.0 (216.0–303.0)
Creatinine	0.90 (0.80–1.10)
Insurance status	
Medicaid	202 (3.7)
Medicare	2662 (48.3)
Free Care	104 (1.9)
Self-pay	73 (1.3)
Zip code characteristics	
Median household income (\$)	83,003 (62,997–102,539)
Median age (y)	42.3 (38.0–45.1)
High school graduation rate (%)	0.23 (0.14–0.30)
Unemployment rate (%)	0.055 (0.045–0.070)
Population density	1924 (766–5766)
Medications	
Angiotensin converting enzyme inhibitor	566 (10.3)
Angiotensin receptor blocker	249 (4.5)
Antidepressant	488 (8.9)
Beta-2-agonist	272 (4.9)
Beta-blocker	689 (12.5)
Benzodiazepines	422 (7.7)
Gabapentin	223 (4.0)
Immunosuppressant	745 (13.5)
Nonsteroidal anti-inflammatory drug	969 (17.6)
Opioid	1213 (22.0)
Antipsychotic	119 (2.2)
Comorbidities	
Tobacco use	246 (4.5)
Alcohol abuse	62 (1.1)
Drug abuse	60 (1.1)
Diabetes	385 (7.0)
Renal failure	136 (2.5)
Depression	511 (9.3)
Psychosis	28 (0.5)
Myocardial infarction	210 (3.8)
Congestive heart failure	216 (3.9)
Peripheral vascular disease	209 (3.8)
Cerebrovascular accident	158 (2.9)
Chronic obstructive pulmonary disease	529 (9.6)
Arrhythmias	739 (13.4)
Valvular disease	303 (5.5)
Liver disease	114 (2.1)
Solid tumor	153 (2.8)
Preoperative opioid duration	
None	4451 (80.8)
<180 d	556 (10.1)
>180 d	500 (9.1)
Sustained opioid prescription	345 (6.3)

IQR, interquartile range.

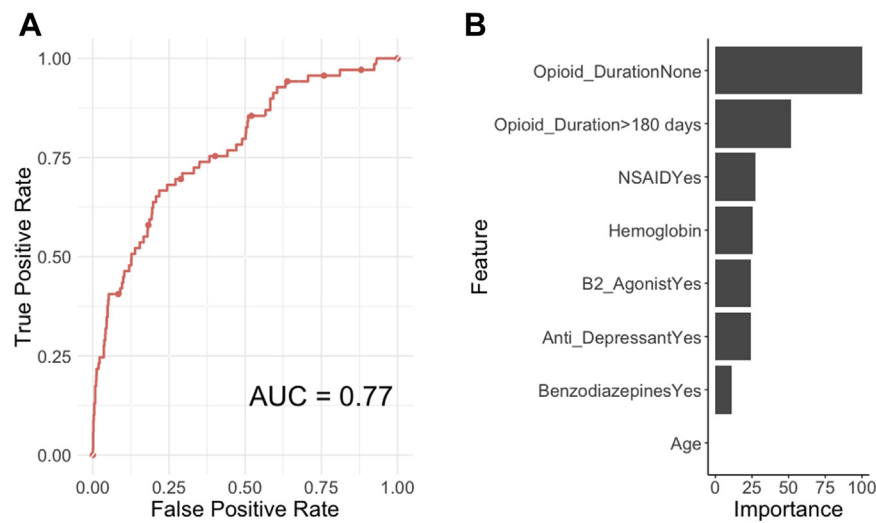


Fig. 1. (A) c-statistic (area under the receiver operating curve) for elastic-net penalized logistic regression. (B) Global variable importance for prediction of opioid dependence. AUC, area under receiver operating curve.

Model Development

Training and testing sets were created with an 80:20 stratified split of the available patient population. The training set was used to develop the algorithms. The subset of variables determined for final modeling was selected by recursive feature elimination with random forest algorithms [20]. Stochastic gradient boosting (SGB), random forest, support vector machine (SVM), neural network, and elastic-net penalized logistic regression (ENPLR) were chosen as candidate models on the basis of prior studies [13,21]. The [Supplementary Methods](#) include a list of other algorithms considered. Technical description of the development process has been described extensively previously [13,22]; in brief, 10-fold cross-validation repeated 3 times was used to develop and assess the models in the training set and the models were assessed by discrimination, calibration, and overall performance (Brier score) [23].

Discrimination was ascertained with the receiver operating curve and the area under the receiver operating curve (AUC), also known as c-statistic for binary classification, was calculated. Perfect models have AUC = 1, whereas models no better than chance have AUC = 0.5. Calibration was ascertained with the calibration plot and the calibration intercept and slope. Calibration refers to the concurrence of the predicted probabilities with the observed frequencies of events in a given population. Perfect models have calibration intercept = 0 and calibration slope = 1. The Brier score is defined as the mean squared difference between the predicted probabilities and the observed outcomes [24]. The incremental benefit of the algorithms was assessed by comparing the Brier score of the algorithms to the null model Brier score, which is defined as the Brier score for the model that assigns a predicted probability equal to the prevalence of the outcome in the study population for

every patient in the population. Decision curve analysis was undertaken to measure the expected utility if management changes based on these models in practice [25]. The decision curve plots net benefit on the y-axis and predicted probabilities on the x-axis. Net benefit is defined as the weighted average of true positives and false positives generated by the model at a given threshold. Because individual clinicians and individual patients have different thresholds for decision making, the decision curve analysis allows users to gauge the expected net benefit of these models at a specific individualized threshold without prescribing a specific management change. Furthermore, the utility of the models is compared to the default strategies of changing management for all patients or for no patients. Final models from the training set were assessed along these dimensions again on the independent testing set not used for model development. Model interpretability and explanation was provided at the global and local levels [26]. Global explanations were provided with variable importance plots. Variable importance plots show the relative importance of variables used for predicted indexed against the most important variable. Finally, local explanations were provided for individual patients to demonstrate which factors for the specific patient in question contributed to the prediction generated by the model [27].

Digital Application

The best performing algorithm was incorporated into a digital application capable of providing both predictions and explanations. Data analysis was conducted with the Anaconda Distribution (Anaconda, Inc, Austin, TX), R version 3.5.1 (The R Foundation, Vienna, Austria), RStudio version 1.1.453 (RStudio, Boston, MA), and Python version 3.6 (Python Software Foundation, Wilmington, DE).

Table 2
Discrimination and Calibration of Algorithms on Cross-Validation of Training Set, n = 4406.

Metric	Stochastic Gradient Boosting	Random Forest	Support Vector Machine	Neural Network	Elastic-Net Penalized Logistic Regression
AUC	0.81 (0.79-0.83)	0.79 (0.75-0.79)	0.61 (0.58-0.63)	0.81 (0.79-0.83)	0.81 (0.79-0.83)
Intercept	0.07 (−0.10 to 0.23)	0.04 (−0.24 to 0.31)	0.67 (−0.05 to 1.39)	0.11 (−0.07 to 0.28)	0.18 (−0.01 to 0.36)
Slope	1.03 (0.96-1.10)	0.58 (0.49-0.67)	1.24 (0.98-1.51)	1.09 (1.01-1.17)	1.09 (1.00-1.17)
Brier	0.051 (0.049-0.052)	0.055 (0.054-0.057)	0.058 (0.057-0.059)	0.051 (0.050-0.053)	0.051 (0.049-0.053)

Data are expressed as mean (95% confidence interval). Null model Brier score = 0.059. AUC, area under the receiver operating curve.

Table 3
Discrimination and Calibration of Algorithms in Holdout Set, n = 1101.

Metric	Stochastic Gradient Boosting	Random Forest	Support Vector Machine	Neural Network	Elastic-Net Penalized Logistic Regression
AUC	0.77	0.75	0.59	0.77	0.77
Intercept	−0.20	0.09	1.14	−0.15	0.01
Slope	0.87	0.61	1.42	1.09	0.97
Brier	0.053	0.056	0.055	0.053	0.052

Null model Brier score = 0.059.

AUC, area under the receiver operating curve.

Results

Overall, 5507 patients underwent THA for osteoarthritis with 345 (6.3%) demonstrating prolonged postoperative opioid prescriptions. The median age was 66 years (interquartile range 57–74) and 2725 (49.5%) patients were female (Table 1). Before surgery, 556 (10.1%) patients had opioid prescriptions for less than 180 days and 500 (9.1%) had opioid prescriptions for 180 days or more.

Variables determined for prediction of prolonged opioid prescriptions were age, duration of opioid exposure, preoperative hemoglobin, and preoperative medications (antidepressants, benzodiazepines, nonsteroidal anti-inflammatory drugs, and beta-2-agonists) (Fig. 1B).

On cross-validation of the training set, n = 4406, the AUC ranged from 0.61 for the SVM to 0.81 for SGB, neural network, and ENPLR (Table 2). The calibration intercept ranged from 0.07 to 0.67 and the calibration slope ranged from 0.58 to 1.24. The null model Brier score in the study population was 0.059 and the algorithms' Brier scores ranged from 0.051 to 0.058. In the testing set, n = 1101, the AUC ranged from 0.59 (SVM) to 0.77 (SGB and ENPLR) (Table 3). The calibration intercept ranged from −0.20 to 1.14 and the calibration slope ranged from 0.61 to 1.42. The Brier score ranged from 0.052 to 0.056. Overall, the ENPLR performed the best on both cross-validation of the training and assessment in the testing set with AUC = 0.77, calibration intercept = 0.01, calibration slope = 0.97, and Brier score = 0.052 (Figs. 1A and 2A). On decision curve

analysis, the model exhibited higher net benefit than the default strategies of changing management for all patients or no patients (Fig. 2B).

An example of a local, individual patient-level explanation for the model predictions is shown in Figure 3. For a 62-year-old patient with preoperative opioid prescriptions for greater than 180 days, antidepressant use, and anemia, the predicted probability of prolonged postoperative opioid prescriptions was 0.32. In this case, duration of opioid use, antidepressant use, and anemia increased the probability of postoperative opioid prescriptions, whereas the lack of other preoperative medication use (benzodiazepines, beta-2-agonist, nonsteroidal anti-inflammatory) and age decreased the probability.

The final model presented as a digital application accessible on desktops, tablets, and smartphones is available at <https://sorg-apps.shinyapps.io/thaopioid/>. Default values in the application are placeholders and the model requires complete information in order to generate predictions and explanations.

Discussion

The national opioid epidemic has increased pressure on orthopedic surgeons to modify opioid prescribing habits. Opioids are very effective for acute pain management after procedures such as THA, but previous studies have highlighted the risk of short-term exposures in the postoperative period leading to prolonged opioid use and dependence. In line with the national data on increasing rates of prescription opioid use, previous studies have shown that opioid use after THA has also increased from 2007 to 2015 [8]. In this context of rising rates of deaths related to prescription opioids and expected increases in THA utilization in the next decade, orthopedic surgeons need tools to effectively manage postoperative pain in THA while reducing the potential for long-term adverse consequences. The ENPLR model developed in this study may be one potential solution for this need with good performance on discrimination, calibration, and decision curve analysis for preoperative prediction of prolonged postoperative opioid prescriptions after THA.

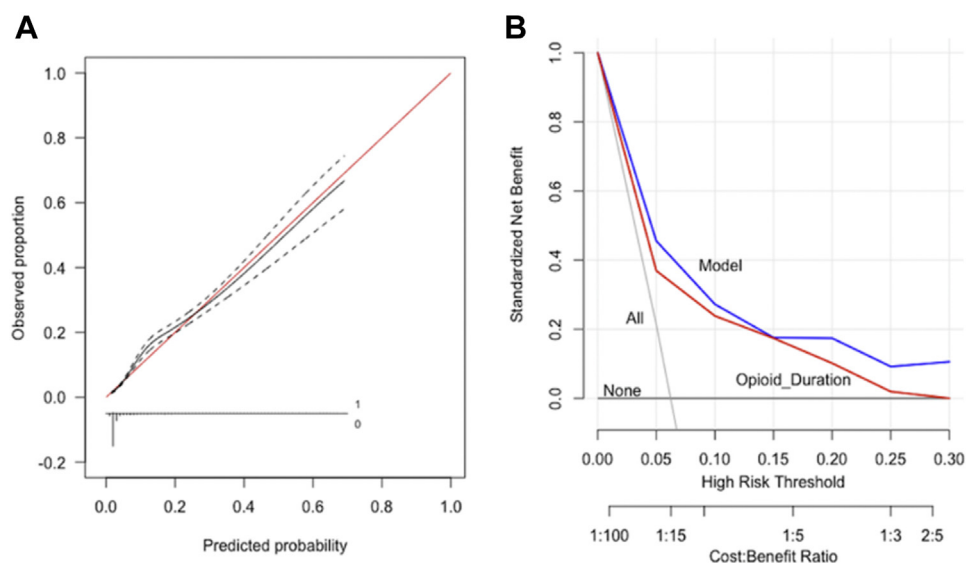


Fig. 2. (A) Calibration plot for elastic-net penalized logistic regression. (B) Decision curve analysis for elastic-net penalized logistic regression showing net benefit of the model relative to default strategies of changing management for all patients or for no patients. The horizontal line “None” represents the default strategy of changing management for no patients (net benefit is zero at all thresholds). The “All” line represents the net benefit from the default strategy of changing management for all patients. The line slopes down because at a threshold of zero, false positives are given no weight relative to true positives; as the threshold increases, false positives gain increased weight relative to true positives and the strategy of changing management for all patients results in decreasing net benefit.

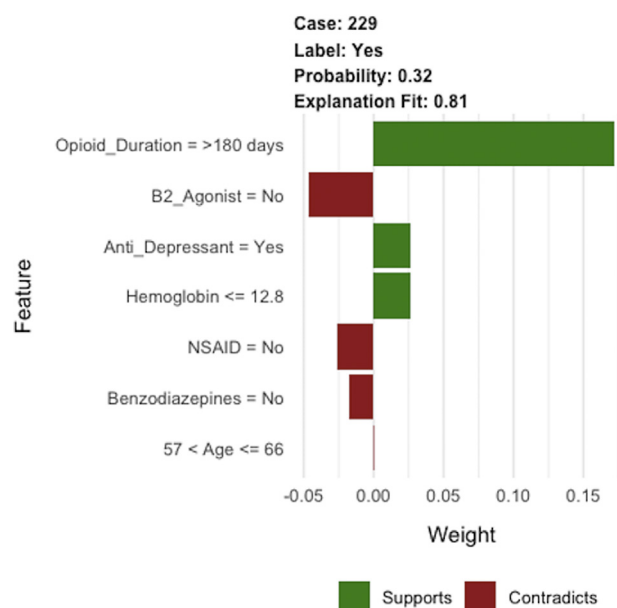


Fig. 3. Example of individual patient-specific explanation generated by the elastic-net penalized logistic regression.

Variables identified in this study for prediction of prolonged postoperative opioid prescriptions concurred with previous studies of opioids. Goesling et al [3] studied 6-month opioid use in 331 THA patients and found that 34.7% of patients with preoperative opioid use continued on opioids at 6 months vs 4.3% of opioid naïve patients. This effect of preoperative opioid exposure on postoperative opioid use has been confirmed by multiple studies [7,8,28]. Bedard et al [8] studied 1-year postoperative opioid use in 37,393 THA patients from the Humana claims database and also identified age and comorbidities such as anxiety/depression. Medications (benzodiazepines, antidepressants) identified in this study for prediction of prolonged postoperative opioid use similarly concurred with prior studies of cardiac, thoracic, abdominal, and pelvic surgery [5]. The mechanistic link of preoperative hemoglobin (anemia) to prolonged postoperative opioid prescription may be due to the summative effects of multiple comorbidities (renal disease, hematologic diseases, malignancy, rheumatologic/immunologic diseases) in creating derangements in normal physiology in the preoperative period and increasing requirements for pain relief in the postoperative period.

To our knowledge, this is the first study to develop a preoperative algorithm for prediction of postoperative opioid use after THA and the first study to use machine learning for this purpose. The performance of the ENPLR on internal validation was verified through a rigorous assessment consistent with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement. Other studies in arthroplasty have developed machine learning models for prediction of outcomes ranging from costs to complications, mortality, and length of stay [11,29,30]. However, the majority of prior studies have relied on discrimination (AUC or *c*-statistic) as the sole measure of model performance, whereas this study achieved good discrimination while verifying performance on calibration and decision curve analysis. In line with the recommendations for development of clinical prediction models, assessments of calibration must be part and parcel of the development process [31].

Despite the advances noted above, this study has several limitations. First, surveillance of postoperative opioid prescriptions was based on electronic health records and pharmacy records at one

healthcare entity and this does not capture patients who obtained opioids from providers outside this entity or through illicit means. Dwyer et al [32] previously followed 186 THA patients in the postoperative period by using a state-wide online prescription registration program; this study was not able to achieve opioid surveillance to this level of detail. Similarly, in the preoperative period, opioid dose in oral morphine equivalents could not be reliably extracted over the full time course of the THA procedures included in this study and the relationship between preoperative opioid exposures was limited to duration of opioid exposure as opposed to both duration and intensity. Furthermore, because the patients included in this study originated from 5 hospitals affiliated with 1 healthcare entity, there is potential for clustering and shared institutional practices that increase the importance of external validation in independent multi-institutional samples to verify the performance of the algorithms developed in this study. Finally, this study is retrospective, observational, and does not account for variations in individual provider prescribing habits which may influence the quantity and duration of postoperative opioid prescriptions after THA.

Acknowledging the limitations noted above, there is a need for future studies to externally validate or refute the findings presented here. With appropriate caution, clinicians may access the digital application provided here and use expertise and judgment to supplement the interpretation of predictions and explanations offered by the final machine learning algorithms. The ultimate purpose of this endeavor is to develop reliable and interpretable algorithms that clinicians can use preoperatively to identify the THA patients at high risk for prolonged postoperative opioid use. Early identification may allow for the pooling of multidisciplinary and institutional resources (mental health, case management, pain service) to reduce the potential for long-term opioid dependence in these patients.

Conclusion

If externally validated in independent populations, the algorithms developed in this study could improve preoperative screening and support for THA patients at high risk for prolonged postoperative opioid prescriptions. Early identification and intervention in high-risk cases may mitigate the long-term adverse consequence of opioid dependence.

References

- [1] Manchikanti L, Singh A. Therapeutic opioids: a ten-year perspective on the complexities and complications of the escalating use, abuse, and nonmedical use of opioids. *Pain Physician* 2008;11(2 Suppl):S63–88.
- [2] Prescription opioid data. In: Atlanta, GA: center for Disease control and prevention; 2018.
- [3] Goesling J, Moser SE, Zaidi B, Hassett AL, Hilliard P, Hallstrom B, et al. Trends and predictors of opioid use after total knee and total hip arthroplasty. *Pain* 2016;157:1259–65.
- [4] Waljee JF, Li L, Brummett CM, Englesbe MJ. Iatrogenic opioid dependence in the United States: are surgeons the gatekeepers? *Ann Surg* 2017;265:728–30.
- [5] Clarke H, Soneji N, Ko DT, Yun L, Wijeyesundera DN. Rates and risk factors for prolonged opioid use after major surgery: population based cohort study. *BMJ* 2014;348:g1251.
- [6] Sun EC, Darnall BD, Baker LC, Mackey S. Incidence of and risk factors for chronic opioid use among opioid-naïve patients in the postoperative period. *JAMA Intern Med* 2016;176:1286–93.
- [7] Inacio MC, Hansen C, Pratt NL, Graves SE, Roughead EE. Risk factors for persistent and new chronic opioid use in patients undergoing total hip arthroplasty: a retrospective cohort study. *BMJ Open* 2016;6:e010664.
- [8] Bedard NA, Pugely AJ, Dowdle SB, Duchman KR, Glass NA, Callaghan JJ. Opioid use following total hip arthroplasty: trends and risk factors for prolonged use. *J Arthroplasty* 2017;32:3675–9.
- [9] Ramkumar PN, Navarro SM, Haerberle HS, Karnuta JM, Mont MA, Iannotti JP, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. *J Arthroplasty* 2019;34:632–7.

- [10] Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol* 2018;6:75.
- [11] Harris AHS, Kuo AC, Weng Y, Trickey AW, Bowe T, Giori NJ. 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;477:452–60.
- [12] Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J Arthroplasty* 2018;33:2358–61.
- [13] Karhade AV, Ogink PT, Thio Q, Broekman MLD, Cha TD, Hershman SH, et al. Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion. *Spine J* 2019;19:976–83.
- [14] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
- [15] Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016;18:e323.
- [16] Brummett CM, Waljee JF, Goesling J, Moser S, Lin P, Englesbe MJ, et al. New persistent opioid use after minor and major surgical procedures in US adults. *JAMA Surg* 2017;152:e170504.
- [17] Oleisky ER, Pennings JS, Hills J, Sivaganesan A, Khan I, Call R, et al. Comparing different chronic preoperative opioid use definitions on outcomes after spine surgery. *Spine J* 2019;19:984–94.
- [18] Schoenfeld AJ, Belmont Jr PJ, Blucher JA, Jiang W, Chaudhary MA, Koehlmoos T, et al. Sustained preoperative opioid use is a predictor of continued use following spine surgery. *J Bone Joint Surg Am* 2018;100:914–21.
- [19] Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2011;28:112–8.
- [20] Kuhn M, Johnson K. *Applied predictive modeling*. New York: Springer; 2013.
- [21] Wainer J. Comparison of 14 different families of classification algorithms on 115 binary datasets. 2019. <https://arxiv.org/abs/1606.00930>.
- [22] Karhade AV, Thio Q, Ogink PT, Bono CM, Ferrone ML, Oh KS, et al. Commentary: Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. *Neurosurgery* 2019.
- [23] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–31, 841–848.
- [24] Brier GW, Allen RA. Verification of weather forecasts. In: *Compendium of meteorology*. 841. Springer; 1951.
- [25] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565.
- [26] Greenwell BM, Boehmke BC, McCarthy AJ. A simple and effective model-based variable importance measure. 2019. <https://arxiv.org/abs/1805.04755>.
- [27] Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning. 2019. <https://arxiv.org/abs/1606.05386>.
- [28] Singh JA, Lewallen D. Predictors of pain and use of pain medications following primary total hip arthroplasty (THA): 5,707 THAs at 2-years and 3,289 THAs at 5-years. *BMC Musculoskelet Disord* 2010;11:90.
- [29] Navarro SM, Wang EY, Haeberle HS, Mont MA, Krebs VE, Patterson BM, et al. Machine learning and primary total knee arthroplasty: patient forecasting for a patient-specific payment model. *J Arthroplasty* 2018;33:3617–23.
- [30] Gabriel RA, Sharma BS, Doan CN, Jiang X, Schmidt UH, Vaida F. A predictive model for determining patients not requiring prolonged hospital length of stay after elective primary total hip arthroplasty. *Anesth Analg* 2019;129:43–50.
- [31] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–31.
- [32] Dwyer MK, Tumpowsky CM, Hiltz NL, Lee J, Healy WL, Bedair HS. Characterization of post-operative opioid use following total joint arthroplasty. *J Arthroplasty* 2018;33:668–72.

Appendix

Supplementary Methods

Stochastic gradient boosting, random forest, support vector machine, neural network, and elastic-net penalized logistic regression were chosen as candidate models on the basis of prior studies [1,2]. Other models considered for inclusion in the study were: (1) logistic regression, (2) radial basis function, (3)

k-nearest neighbor, (4) naïve Bayes, (5) learning vector quantization, (6) linear discriminant analysis, (7) quadratic discriminant analysis, (8) flexible discriminant analysis, (9) generalized additive models, (10) decision trees, and (11) partial least squares [3–5].

Supplementary Table 1
List of Medications by Category.

Group	Medications
ACE inhibitor	Captopril, lisinopril, enalapril, ramipril, benazepril, fosinopril, moexipril, perindopril, quinapril, trandolapril
ARB	Losartan, valsartan, azilsartan, candesartan, eprosartan, irbesartan, olmesartan, telmisartan
Antidepressant	Citalopram, escitalopram, paroxetine, fluoxetine, fluvoxamine, sertraline, desvenlafaxine, duloxetine, levomilnacipran, milnacipran, tofenacin, venlafaxine, vilazodone, vortioxetine, etoperidone, trazodone, nefazodone, reboxetine, viloxazine, amitriptyline, butriptyline, clomipramine, desipramine, dosulepin, imipramine, iprindole, lofepramine, melitracen, nortriptyline, protriptyline, trimipramine, opipramol, tianeptine, amoxapine, maprotiline, mianserin, mirtazapine, setiptiline, isocarboxazid, phenelzine, tranlycypromine, selegiline, metralindole, moclobemide, pirlindole, toloxatone
Beta-2-agonist	Bitolterol, fenoterol, isoproterenol, levalbuterol, pirbuterol, procaterol, albuterol, arformoterol, bambuterol, clenbuterol, formoterol, salmeterol, abediterol, carmoterol, indacaterol, olodaterol, vilanterol
Beta-blocker	Acebutolol, atenolol, betaxolol, bisoprolol, carteolol, carvedilol, esmolol, labetalol, metoprolol, nadolol, nebivolol, penbutolol, pindolol, propranolol, sotalol, timolol
Benzodiazepines	Alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, estazolam, flurazepam, halazepam, lorazepam, oxazepam, prazepam, quazepam, temazepam, triazolam
Immunosuppressant	Mycophenolate, etanercept, adalimumab, azathioprine, cyclosporine, tacrolimus, sirolimus, infliximab, natalizumab, methotrexate, certolizumab, prednisone, dexamethasone, cortisone, hydrocortisone, prednisolone, leflunomide, cyclophosphamide, chlorambucil, budesonide, everolimus, abatacept, basiliximab, daclizumab, muromonab, vedolizumab, ustekinumab, tocilizumab, secukinumab, rituximab, ixekizumab, golimumab, anakinra
NSAID	Celecoxib, diclofenac, diflunisal, etodolac, flurbiprofen, ibuprofen, ketoprofen, meloxicam, nabumetone, naproxen, piroxicam, sulindac, ibuprofen, indomethacin, ketorolac
Opioid	Hydromorphone, oxycodone, hydrocodone, fentanyl, methadone, morphine, codeine, butorphanol, nalbuphine, tramadol, tapentadol
Typical antipsychotic	Chlorpromazine, thioridazine, mesoridazine, levomepromazine, loxapine, molindone, perphenazine, thiothixene, trifluoperazine, haloperidol, fluphenazine, droperidol, zuclopenthixol, prochlorperazine
Atypical antipsychotic	Amisulpride, aripiprazole, asenapine, blonanserine, clotiapine, clozapine, iloperidone, lurasidone, mosapramine, olanzapine, paliperidone, perospirone, quetiapine, remoxipride, risperidone, sertindole, sulpiride, ziprasidone, zotepine

ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; NSAID, non-steroidal anti-inflammatory drugs.

References

[1] Karhade AV, Ogink PT, Thio Q, Broekman MLD, Cha TD, Hershman SH, et al. Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion. *Spine J* 2019;19:976.

[2] Wainer J. Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv preprint arXiv:1606.00930*. 2016.

[3] Kuhn M, Johnson K. *Applied predictive modeling*. Springer; 2013.

[4] Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning: Springer series in statistics* New York. 2001.

[5] James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning*. Springer; 2013.