European Journal of Pain

EJP European Journal of Pain

ORIGINAL ARTICLE OPEN ACCESS

Development and Internal Validation of a Machine Learning Model for Predicting Long-Term Opioid Therapy After Hip Fracture Surgery in Older, Opioid-Naïve Adults

Yasmina Maria Tudorache[1,2] | Simon Storgaard Jensen[1,2] | Katie Jane Sheehan[3] | Jan-Erik Gjertsen[4,5] | Alma Becic Pedersen[1,2]

[1]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark | [2]Department of Clinical Medicine, Aarhus University Hospital, Aarhus, Denmark | [3]Center of Bone and Joint Health, Blizard Institute, Queen Mary, University of London, London, UK | [4]Department of Orthopaedic Surgery, Haukeland University Hospital, Bergen, Norway | [5]Department of Clinical Medicine, University of Bergen, Bergen, Norway

Correspondence: Alma Becic Pedersen (abp@clin.au.dk)

Received: 29 July 2025 | **Revised:** 14 April 2026 | **Accepted:** 16 April 2026

Keywords: hip fracture | long-term opioid usage | opioid-naïve | pain management | prediction model

ABSTRACT

Background: Long-term opioid therapy (LTOT) after hip fracture surgery is a common postoperative complication associated with adverse outcomes, yet tools to identify at-risk patients among opioid-naïve older adults are lacking. This study aimed to develop and internally validate a parsimonious model to predict LTOT following hip fracture surgery.

Methods: Using Danish nationwide registries, we identified 26,057 opioid-naïve patients (≥65 years) undergoing hip fracture surgery (2010–2020) and analysed 29 predictors, covering demographics, comorbidities, medication, lifestyle, socioeconomics and surgical factors. LTOT was defined as redeeming ≥2 prescriptions between 31 and 365 days of surgery. Models were developed using four machine learning methods: logistic regression, backwards elimination, elastic net penalised logistic regression and random forest. Model performance was assessed using area under the receiver operating characteristic curve (AUC), calibration slope, intercept, Brier score and decision curve analysis.

Results: LTOT was identified in 8095 (31.1%) patients. The backward elimination algorithm identified the best performing model, selecting 8 of 29 predictors and achieving an AUC of 0.68, calibration slope of 0.99, intercept of 0.02 and Brier score of 0.20. Predictors included age, marital status, preoperative non-opioid pain medication, preoperative novel oral anticoagulants, fracture type, surgery delay, length of hospital stay and postoperative cumulated ambulation score at discharge.

Conclusions: A prediction model was developed and validated for use at discharge to identify patients at risk of LTOT 1 year after hip fracture. The model may support risk stratification at discharge, but requires external validation and evaluation of clinical implementation before routine use.

Significance Statement: This study presents the first internally validated prediction model for long-term opioid use in opioid-naïve older adults after hip fracture surgery. The model functions as a simple and interpretable risk stratification tool at discharge and has been incorporated in a freely available risk calculator. It addresses the lack of clinically applicable risk stratification tools in this frail population and highlights opportunities for more targeted postoperative pain management, although feasibility testing is required before clinical implementation.

1 | Introduction

The global burden of hip fractures is expected to increase due to the aging population (Kristensen et al. 2020; Sing et al. 2023). Management of a hip fracture usually involves a combination of urgent surgery (Bhandari and Swiontkowski 2017), rehabilitation and medication, in which opioids play an important role in addressing short-term postoperative pain (Simoni et al. 2019). Despite advancements in peri- and post-operative care including pain management (Chou et al. 2016; Kehlet 2020; Mazarello Paes et al. 2025; Tudorache et al. 2025), many hip fracture patients rely on opioid prescriptions for pain relief during their recovery (Chou et al. 2016; Simoni et al. 2019). 17% to 27% of opioid-naïve hip fracture patients continue opioid use 1 year after surgery (Risbo et al. 2025; Simoni et al. 2019).

Opioid usage increases the risk of falls, which can lead to subsequent fractures (Yue et al. 2020) and elevated mortality (Lindestrand et al. 2015). Moreover, complications such as respiratory depression, dizziness and postoperative delirium (Benyamin et al. 2008) as well as adverse drug interactions (Matos et al. 2020) can impede recovery and increase healthcare costs (Johnson et al. 2022).

Preoperative opioid use is a consistent predictor of long-term opioid therapy (LTOT) after hip fracture surgery (Chaudhary et al. 2019; Edwards et al. 2021; Hereford et al. 2022; Stone et al. 2023), but its strong effect may overshadow other important predictors, particularly those relevant to identifying risk among opioid-naïve patients. The complex interplay of risk factors makes individualised risk assessment difficult. Predictive models for LTOT after elective hip arthroplasty exist (Karhade et al. 2019; Quaye et al. 2025), but lack universal clinical uptake. Furthermore, their applicability to acute hip fracture patients remains uncertain, as these cases present a different clinical entity than elective patients. Therefore, to address this knowledge gap, this study aimed to develop and validate a tool to identify opioid-naive hip fracture patients at high risk of LTOT.

2 | Materials and Methods

2.1 | Guidelines

This study is reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis + Artificial Intelligence guidelines (Collins et al. 2024) and follows best practice guidance for model development and validation by Steyerberg and Vergouwe (2014).

2.2 | Data Sources

This study utilised data from five Danish population-based nationwide medical databases. Healthcare in Denmark is universally accessible and funded through taxation. Databases were linked using the unique 10-digit civil registration number assigned through the Civil Registration System, which tracks civil, vital and migration status (Schmidt et al. 2014). This system allowed for the identification and inclusion of all eligible patients who had undergone hip fracture surgery. Additional description of the five databases is found in the Supporting Information S1.

2.3 | Study Population

Opioid-naïve patients with a femoral neck or per/subtrochanteric fracture recorded in the Danish Multidisciplinary Hip Fracture Registry database between 01.01.2010 and 31.12.2020 were included. Those with a record of opioid use in the 12-months pre-fracture and/or who had an inactive civil registration number before the postoperative follow-up period between 31 and 365 days of surgery began were excluded, see Figure 1 and Figure S1 for graphical depictions.

2.4 | Outcome

LTOT was defined as a patient redeeming ≥2 opioid prescriptions from a community pharmacy between 31 and 365 days of surgery. The included opioids and corresponding ATC codes are listed in Table S1.

Opioids dispensed within the first 30 days postoperatively were excluded, as these were assumed to reflect acute postoperative pain and not long-term use.

Data from patients who died or emigrated (inactive civil registration number) before the end of the 12-month period were retained in the outcome assessment to reduce the risk of immortal time bias in model development (Reps et al. 2021). Patients

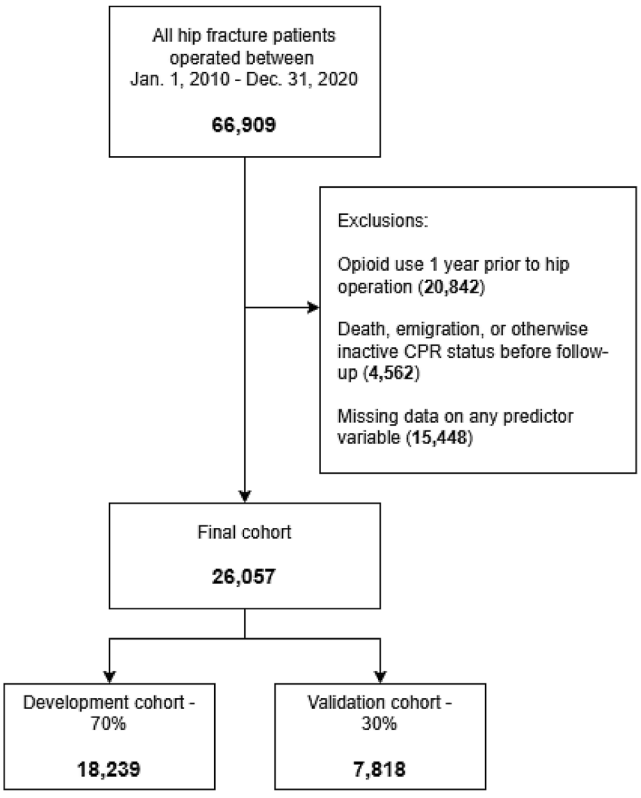

FIGURE 1 | Study population flowchart.

who died/emigrated before redeeming two prescriptions were classified as non-LTOT users, while those who redeemed two prescriptions were classified as LTOT users.

2.5 | Predictors

The criteria for selecting candidate predictors included a documented or plausible association with LTOT following hip fracture (Edwards et al. 2021; Gadgaard et al. 2024; Karhade et al. 2019), as well as the ability to be easily obtained at the time of admission or at discharge. We identified 29 candidate predictors through our databases: age and gender; preoperative sociodemographic factors including education level, living situation and marital status; preoperative clinical predictors, including surgical delay > 24 h, body mass index (BMI), cumulated ambulation score (CAS), hip fracture type, prior prescriptions across nine different medication groups (Edwards et al. 2021) and the presence of major comorbidities commonly found in a hip fracture population (Gadgaard et al. 2024). Medication use was assessed during the 6 months preceding the hip fracture operation date, while comorbidity status was assessed 5 years prior to the operation. Lastly, we included postoperative factors, such as postoperative CAS, pre-to-postoperative CAS change and hospital stay > 8 days (corresponding to the median length of hospitalisation in the cohort). The individual predictor levels are presented in Table 1. Codes for the included predictors are in Tables S2–S4. All predictors were coded as categorical variables for ease of implementation in the machine learning models and clinical practice.

2.6 | Missing Data

Patients with missing data on any predictor variable were excluded from model development and validation. A complete-case approach was chosen due to the moderate proportion of missingness, the large sample size and prior studies using this dataset demonstrating negligible differences between multiple imputation and complete-case analyses for similar predictors (Hjelholt et al. 2022; Pedersen et al. 2017).

2.7 | Analytical Methods

The selection of modelling approaches was guided by prior prediction studies (Heinze et al. 2018; Karhade et al. 2020, 2019; Katakam et al. 2020) and by the structure of our dataset, which included a moderate number of clinically defined predictors from structured electronic health records. Because the events-per-variable ratio was high and complex machine learning models have not consistently outperformed traditional approaches (Christodoulou et al. 2019; Lynam et al. 2020; Sun et al. 2022), we focused on regression and tree-based methods that balance predictive performance, interpretability and clinical applicability.

Patients were followed from postoperative Day 31 until death, emigration, or 365 days after surgery, whichever occurred first. Patient characteristics were summarised using descriptive statistics.

2.8 | Model Development

The cohort was randomly divided into a development set (70%) and a validation set (30%), with a fixed seed applied to ensure reproducibility. The development set was used to train the machine learning models, while the validation set—containing data not seen during model training—was used to evaluate their performance.

Logistic regression, backwards elimination, elastic-net penalised logistic regression and random forest were chosen as candidate machine learning models based on results from previously published literature (Heinze et al. 2018; Karhade et al. 2019; Steyerberg 2019). The logistic regression model was implemented as a standard multivariable logistic regression. For backwards elimination, variable selection was performed across 500 bootstrap samples and predictors retained in at least 50% of the samples were included in the final model, which was a multivariable logistic regression model (Heinze et al. 2018; Steyerberg 2019). Both the elastic-net penalised logistic regression and the random forest models were tuned using 10-fold cross-validation repeated three times, with the area under the receiver operating characteristic curve (AUC) value used as the optimisation metric (Karhade et al. 2019). Additional technical details are provided in the Table S5.

2.9 | Model Performance

To assess the discrimination and calibration of our models, we evaluated their performance based on the AUC, calibration intercept and slope, Brier score and null model Brier score on the validation set (Steyerberg and Vergouwe 2014). Perfect models have an AUC of 1, whereas uninformative random models have an AUC of 0.5 (Steyerberg 2019). Calibration assesses how closely predicted probabilities agree with observed outcomes (Steyerberg 2019; Steyerberg and Vergouwe 2014). Calibration intercept of 0 and slope of 1 characterise a hypothetical model with perfect predictions (Steyerberg and Vergouwe 2014). The Brier score is defined as the squared difference between the observed outcomes and predicted probabilities (Steyerberg 2019), with lower scores indicating better predictions. The improvement in predictions provided by the algorithms was evaluated by comparing their Brier scores to that of a null model, which assigns every patient a predicted probability equal to the outcome's prevalence in the study population (Karhade et al. 2019). The best performing model based on performance metrics was further validated using decision curve analysis.

2.10 | Decision Curve Analysis

The expected clinical utility of the best-performing prediction model was evaluated using decision curve analysis, a method that quantifies the net benefit of using a model to guide interventions compared to treating all or no patients, across a range of risk thresholds (Kerr et al. 2016; Vickers and Elkin 2006).

TABLE 1 | Baseline patient characteristics.

	All patients	No LTOT	LTOT
	26,057 (100.00%)	17,962 (100.00%)	8095 (100.00%)
Variable			
Sex			
Female	17,743 (68.1%)	12,192 (67.9%)	5551 (68.6%)
Male	8314 (31.9%)	5770 (32.1%)	2544 (31.4%)
Age category			
65–74	6405 (24.6%)	4283 (23.8%)	2122 (26.2%)
75–84	10,509 (40.3%)	7307 (40.7%)	3202 (39.6%)
85+	9143 (35.1%)	6372 (35.5%)	2771 (34.2%)
BMI category			
BMI < 18.5	2154 (8.3%)	1535 (8.6%)	619 (7.7%)
BMI 18.5–24.9	14,112 (54.2%)	9760 (54.3%)	4352 (53.8%)
BMI 25–29.9	7521 (28.9%)	5141 (28.6%)	2380 (29.4%)
BMI ≥ 30	2270 (8.7%)	1526 (8.5%)	744 (9.2%)
Fracture type			
Femoral neck	15,185 (58.3%)	10,850 (60.4%)	4335 (53.6%)
Per/subtrochanteric	10,872 (41.7%)	7112 (39.6%)	3760 (46.5%)
Preoperative CAS			
0–5	2763 (10.6%)	1847 (10.3%)	916 (11.3%)
6	23,294 (89.4%)	16,115 (89.7%)	7179 (88.7%)
Postoperative CAS			
0–5	15,611 (59.9%)	10,143 (56.5%)	5468 (67.6%)
6	10,446 (40.1%)	7819 (43.5%)	2627 (32.5%)
Change in CAS			
Any increase	656 (2.5%)	465 (2.6%)	191 (2.4%)
No change (0)	10,880 (41.8%)	8096 (45.1%)	2784 (34.4%)
Decreased 1–2 points	6070 (23.3%)	4086 (22.8%)	1984 (24.5%)
Decreased 3–4 points	7750 (29.7%)	4851 (27.0%)	2899 (35.8%)
Decreased 5–6 points	701 (2.7%)	464 (2.6%)	237 (2.9%)
Surgical delay > 24 h	8063 (30.9%)	5854 (32.6%)	2209 (27.3%)
Hospital stay > 8 days	7719 (29.6%)	6555 (36.5%)	1164 (14.4%)
Education level			
Low	14,165 (54.4%)	9885 (55.0%)	4280 (52.9%)
Medium	8619 (33.1%)	5871 (32.7%)	2748 (34.0%)
High	3273 (12.6%)	2206 (12.3%)	1067 (13.2%)
Living situation			
Alone in own home	11,163 (42.8%)	7909 (44.0%)	3254 (40.2%)
Cohabiting in own home	8985 (34.5%)	6365 (35.4%)	2620 (32.4%)
Nursing home	4441 (17.0%)	2644 (14.7%)	1797 (22.2%)
Missing, homeless or other	1468 (5.6%)	1044 (5.8%)	424 (5.2%)

(Continues)

TABLE 1 | (Continued)

	All patients	No LTOT	LTOT
Marital status			
Married	9406 (36.1%)	6600 (36.7%)	2806 (34.7%)
Never married	1730 (6.6%)	1107 (6.2%)	623 (7.7%)
Divorced	3178 (12.2%)	2076 (11.6%)	1102 (13.6%)
Widowed	11,743 (45.1%)	8179 (45.5%)	3564 (44.0%)
Preoperative comorbidities			
Cardiovascular disease	10,797 (41.4%)	7476 (41.6%)	3321 (41.0%)
Pulmonary disease	2504 (9.6%)	1764 (9.8%)	740 (9.1%)
Renal or haematological disease	4006 (15.4%)	2837 (15.8%)	1169 (14.4%)
Hepatic or gastric disease	748 (2.9%)	524 (2.9%)	224 (2.8%)
Metabolic disease	3930 (15.1%)	2712 (15.1%)	1218 (15.1%)
Musculoskeletal disease	867 (3.3%)	581 (3.2%)	286 (3.5%)
Malignant disease	3164 (12.1%)	2199 (12.2%)	965 (11.9%)
Neurological or psychiatric disease	5745 (22.1%)	3744 (20.8%)	2001 (24.7%)
Preoperative medications			
Non-opioid pain medication	9271 (35.6%)	6097 (33.9%)	3174 (39.2%)
Psychiatric medication	8571 (32.9%)	5836 (32.5%)	2735 (33.8%)
NOAC medication	1528 (5.9%)	906 (5.0%)	622 (7.7%)
Antiplatelets medication	7613 (29.2%)	5423 (30.2%)	2190 (27.1%)
Vitamin K antagonists	1757 (6.7%)	1294 (7.2%)	463 (5.7%)
Antibiotics	7175 (27.5%)	4921 (27.4%)	2254 (27.8%)
Antiosteoporosis medication	2151 (8.3%)	1429 (8.0%)	722 (8.9%)
Statins	7214 (27.7%)	4963 (27.6%)	2251 (27.8%)
Corticosteroids	1393 (5.4%)	966 (5.4%)	427 (5.3%)
Death during follow up	3931 (15.1%)	2781 (15.5%)	1150 (14.2%)

Note: Numbers are shown as counts and percentages.
Abbreviations: BMI, body mass index; CAS, cumulated ambulation score; h, hours; LTOT, long-term opioid therapy; NOAC, novel oral anticoagulants.

2.11 | Model Access

The final model is accessible as a freely available Excel spreadsheet (Supporting Information S1—Opioid Prediction Model).

2.12 | Statistics Software

All analyses were done using R version 4.4.1 and RStudio version 2023.12.1.0 on the remote servers of Statistics Denmark (Table S5).

2.13 | Ethical Approval

This study is reported to the Danish Protection Agency through registration at Aarhus University (record number: AU-2016-051-000001, sequential number 880). Since studies are based on existing Danish registry data and no patient interventions will take place, ethical approval was not required.

3 | Results

3.1 | Population Description

Of 26,057 opioid-naïve adults with complete predictor data who underwent hip fracture surgery during 2010–2020, 8095 (31.1%) met the definition for LTOT. Further population descriptive statistics can be seen in Table 1.

3.2 | Model Performance

The backwards elimination model identified eight predictors: age, fracture type, postoperative CAS, marital status, preoperative non-opioid pain medication use, preoperative novel oral anticoagulant (NOAC) use, surgical delay > 24 h and hospital stay > 8 days (Table 2). Graphical depictions of the distribution of these predictors by LTOT outcome are shown in Figure S2. These variables were included in a multivariable logistic regression model, which served as the final prediction

TABLE 2 | Odds ratios and predictive model coefficients for the backwards elimination model.

Variable	Odds ratio (95% CI)	Model beta coefficient
Intercept		−0.929
Fracture type		
Femoral neck	Ref.	
Per/subtrochanteric	1.33 (1.24, 1.42)	0.283
Postoperative CAS		
6	Ref.	
0–5	1.77 (1.65, 1.91)	0.573
Marital status		
Married	Ref.	
Divorced	1.32 (1.18, 1.47)	0.276
Widowed	1.08 (1, 1.17)	0.081
Never married	1.34 (1.17, 1.53)	0.291
Preoperative non-opioid pain medication		
No	Ref.	
Yes	1.22 (1.14, 1.31)	0.200
Preoperative NOAC medication		
No	Ref.	
Yes	1.55 (1.35, 1.77)	0.437
Surgical delay > 24 h		
No	Ref.	
Yes	0.8 (0.75, 0.86)	−0.219
Hospital stay > 8 days		
No	Ref.	
Yes	0.27 (0.25, 0.29)	−1.312
Age category		
65–74	Ref.	
75–84	0.84 (0.77, 0.92)	−0.174
85+	0.78 (0.71, 0.86)	−0.249

Abbreviations: 95 CI, 95% confidence intervals; CAS, cumulated ambulation score; CI, confidence intervals; NOAC, novel oral anticoagulants.

model based on the backwards elimination algorithm. This model demonstrated the best performance in both discrimination (AUC 0.68; 95% CI: 0.66–0.69) and calibration (intercept 0.02; slope 0.99) among the four models assessed (Table 3, Figure S3).

Visual inspection of the calibration curve suggested overall acceptable calibration, with some overestimation of predicted risk probabilities beyond the 0.55 threshold (Figure 2). The Brier score for the backwards elimination model was 0.20. For reference, the Brier score of the null model in this population was 0.22, and all tested models yielded Brier scores lower than this baseline, indicating added predictive value (Table 3). The full logistic regression model demonstrated similar performance but required all 29 predictors (Table 3, Figure S3).

TABLE 3 | Discrimination and calibration of the models in the validation set, $n = 7818$.

Metric	Logistic regression	Backwards elimination	ENPLR	Random forest
AUC	0.68	0.68	0.68	0.64
Intercept	0.02	0.02	0.02	0.35
Slope	0.99	0.99	1.06	0.42
Brier	0.20	0.20	0.20	0.21

Note: Null model Brier score = 0.22.
Abbreviations: AUC, area under the receiver operating curve; ENPLR, elastic-net penalised logistic regression.

3.3 | Decision Curve Analysis

The backwards elimination model showed net benefit across risk thresholds for LTOT ranging from 15% to 50% (Figure 3).

To facilitate clinical interpretation, model performance was further evaluated at selected decision thresholds within this range (20%, 30% and 40%), as well as at the optimal threshold based on Youden's J index (32.3%). Sensitivity ranged from 0.45 to 0.89 and specificity from 0.30 to 0.77 across these thresholds. Positive predictive values ranged from 0.37 to 0.47, while negative predictive values ranged from 0.75 to 0.86 (Table 4). For example, if we decide that patients with a risk of 40% or more would be classified as LTOT users, then among 100 patients classified as high risk by the model, 47 would actually develop LTOT (positive predictive value) and among 100 patients classified as low risk, 75 would not become long-term users (negative predictive value). The relatively high negative predictive value indicates that the model is more effective at identifying patients who are unlikely to become long-term users when classified as low risk, which could help clinicians focus follow-up efforts on those at higher risk.

3.4 | Missing Data

Among the 15,448 patients with missing data on any predictor, 7648 (49.5%) were missing BMI, 7518 (48.7%) preoperative CAS, 6504 (42.1%) postoperative CAS and 3308 (21.4%) education level (Table S6). This was primarily due to the fact that registration of these variables was not consistently mandatory in the Danish Multidisciplinary Hip Fracture Registry throughout the entire study inclusion period.

Compared with complete cases, patients with missing data were more often aged ≥ 85 years and more frequently classified as 'missing, homeless, or other' in the living situation variable. Mortality during follow-up was higher in the missing data group (20.7%) than among complete cases (15.1%). The proportion meeting the definition of LTOT was 24.9% among patients with missing predictor data compared with 31.1% in the complete-case cohort.

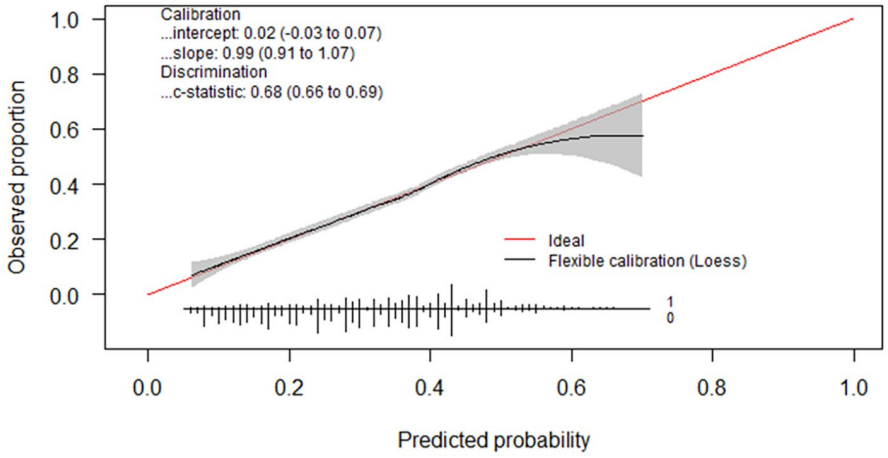

FIGURE 2 | Calibration curve for the backwards elimination model. The C-statistic value corresponds to the model's AUC (area under the curve) value.

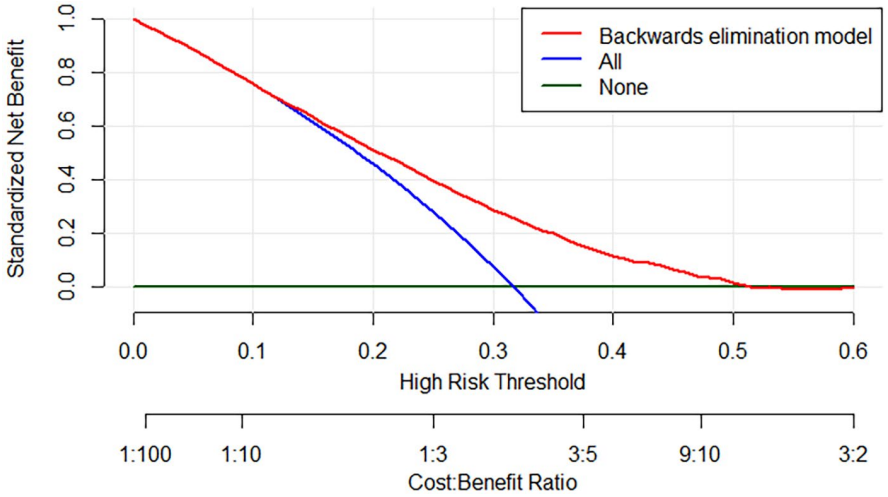

FIGURE 3 | Decision curve analysis of the backwards elimination model in the validation sample. The *x*-axis represents the risk threshold used to define patients at high-risk for long-term opioid therapy. The *y*-axis represents the standardized net benefit (SNB), which can take a value between 0 (no benefit) and 1 (maximal achievable benefit). The blue line represents the hypothetical 'treat all' patients' policy, the green line represents the hypothetical 'treat none' patients policy, the red line represents treating patients based on the backwards elimination model risk profile. The backwards elimination model shows clinical benefit as long as the red line is above both the blue and green lines, which in this case means along risk thresholds between 15% and 50%.

TABLE 4 | Diagnostic performance of the prediction model at selected risk thresholds for LTOT.

Risk threshold	Sensitivity	Specificity	PPV	NPV
20%	0.892	0.296	0.369	0.855
30%	0.716	0.537	0.417	0.803
40%	0.450	0.767	0.473	0.751
32.2%[a]	0.666	0.593	0.432	0.794

Abbreviations: LTOT, long-term opioid therapy; NPV, negative predictive value; PPV, positive predictive value.
[a]Based on Youden's J index.

Despite these differences, opioid prescription patterns were similar between groups, with morphine, oxycodone and tramadol accounting for 96.2% and 97.3% of redeemed prescriptions in the missing data and complete-case groups, respectively.

4 | Discussion

4.1 | Key Findings

In a cohort of 26,057 opioid-naïve adults undergoing hip fracture surgery, 8095 (31.1%) were classified as LTOT users. To predict postoperative LTOT in this patient population, we developed and internally validated four supervised machine learning models. Among these, the backwards elimination model, which retained only 8 of the 29 candidate predictors, demonstrated the best balance between simplicity and predictive performance. Decision curve analysis indicated clinical utility across risk thresholds of 15%–50%. Since the prevalence of LTOT in the study population (31.1%) falls within this range,

the model may offer clinically useful guidance for identifying high-risk patients in real-world settings. This model may support risk stratification for LTOT at discharge, informing individualised postoperative pain management interventions and opioid tapering strategies.

4.2 | Comparison With Other Studies

Few studies have developed prediction models for LTOT in opioid-naïve patients (Karhade et al. 2020; Quaye et al. 2025) and none have specifically targeted hip fracture patients. Comparing our results is further complicated by the lack of consensus on the definition of LTOT (Karmali et al. 2020), which leads to variability in reported prevalence.

Candidate predictors were selected based on previously reported or plausible associations with LTOT after hip fracture surgery (Edwards et al. 2021; Gadgaard et al. 2024; Karhade et al. 2019), as well as their availability in the registries and feasibility for clinical use at discharge. Variables identified in prior studies were therefore considered among the candidate predictors. The final set of predictors was subsequently determined using a data-driven variable selection procedure. Predictors not retained in the final model were excluded based on model-building criteria, including their contribution to optimising model discrimination (AUC) and maintaining a parsimonious and clinically interpretable model. Thus, predictors identified in previous research but not included in our final model were not excluded a priori; rather, they were not retained because they did not contribute sufficiently under the prespecified model-building criteria. This may indicate limited additional predictive value in this dataset (or setting), overlap with other included predictors, or exclusion as part of balancing model discrimination with parsimony and clinical interpretability.

This study identifies a distinct set of predictors for LTOT compared to elective total hip or joint arthroplasty studies. While Karhade et al. (2020, 2019) and Quaye et al. (2025) linked preoperative benzodiazepines, antidepressants and antipsychotic use to LTOT in both opioid-naïve and prior users, our backwards elimination algorithm did not retain these medications, which were grouped under the preoperative psychiatric medication category. However, preoperative gabapentin and non-steroidal anti-inflammatory (NSAID) use, previously identified as predictors (Karhade et al. 2020, 2019; Quaye et al. 2025), were also captured in our model under the broader category of preoperative non-opioid pain relief medications. Edwards et al. (2021) further supports our findings, demonstrating that, among various preoperative medications, NSAID use had the strongest association with LTOT in opioid-naïve patients.

In contrast to previous studies by Quaye et al. (2025) and Chaudhary et al. (2019), which identified increased hospitalisation length as a risk factor for LTOT in elective joint arthroplasty and broader surgical populations, our study found that hospital stays > 8 days and surgical delays > 24 h were associated with a decreased risk. This discrepancy may stem from methodological differences, as the prior studies (Chaudhary et al. 2019; Quaye et al. 2025) excluded patients who died during follow-up, whereas we included them to enhance clinical applicability and generalisability. This approach ensures the model remains relevant for guiding postoperative pain management in the entire hip fracture population, not just those who live through the follow-up period.

While predictor estimates in a prediction study cannot establish causation (van Diepen et al. 2017), we recognise that this inverse association is unexpected. One possible explanation is that longer hospital stays and surgical delays may reflect more complicated clinical cases, where patients die soon after discharge and therefore never become LTOT users. Alternatively, surgical delays might signal preoperative optimisation aimed at better recovery outcomes, while extended hospitalisation could allow closer monitoring and gradual opioid tapering. This idea is supported by Cupp et al. (2023), who found lower LTOT rates among hip fracture patients discharged to skilled nursing facilities, where monitoring is typically more structured, compared to those sent to inpatient rehabilitation facilities.

Though these theories provide insight into potential mechanisms, we chose not to investigate them further, as causal interpretations fall outside the scope of prediction studies.

Interestingly, our model identified marital status as a predictor of LTOT, with widowed, divorced, or never-married patients at higher risk than married patients. This suggests that lack of social support networks may contribute to opioid dependence (Che et al. 2018; Risbo et al. 2025).

4.3 | Strengths and Limitations

The nationwide cohort, complete follow-up, and broad range of tested potential predictors, including socioeconomic factors, strengthen the model's generalisability. Additionally, the use of simple modelling strategies ensures that the prediction model remains easily interpretable for clinicians. However, several limitations must be acknowledged.

Patients with missing predictor data (BMI, pre- and postoperative CAS, education) were excluded (Table S6), representing 37.2% (15,448/41,505) of the opioid-naïve population reachable at follow-up start. Although complete-case analysis may introduce selection bias, prior research using this dataset found that results from complete-case and multiple imputation analyses were similar, indicating minimal impact on estimates (Hjelholt et al. 2022; Pedersen et al. 2017). The study also had a high events-per-variable ratio (~365), exceeding the recommended minimum of 25, thus reducing the risk of model overfitting and enhancing estimate reliability (Heinze et al. 2018).

In line with previously published orthopaedic prediction models (Karhade et al. 2020, 2019; Katakam et al. 2020), we focused on developing a parsimonious and clinically relevant model that could be applied in routine practice. An important aim was to achieve reasonable predictive performance using as few readily available clinical variables as possible. Although several studies have shown that more complex machine learning methods do not consistently outperform regression models (Christodoulou et al. 2019; Lynam et al. 2020; Sun et al. 2022), we acknowledge

that alternative algorithms could potentially yield different results.

Opioid prescription redemption may not accurately reflect actual consumption, as patient adherence and specific clinical indications remain unknown. This could overestimate the actual opioid usage. Nevertheless, prescription opioid dispensing data is a more reliable measure of medication use than most alternatives (Schneeweiss and Avorn 2005). Lastly, unmeasured factors, such as clinician prescribing habits, may influence opioid use.

Regarding additional predictors, pain scores would have been relevant to include; however, our databases lack routinely collected pre- and postoperative pain scores.

4.4 | Clinical Relevance and Future Directions

Implementing the model in an online Excel risk calculator (Supporting Information S1—Opioid Prediction Model) could facilitate exploratory use by healthcare professionals to estimate the risk of LTOT at discharge. This could support tailored postoperative tapering plans and optimise resource allocation by directing preventive interventions toward patients at highest risk, rather than applying them universally. Feasible interventions may include scheduled follow-up with a general practitioner or home nurse visits to provide guidance on opioid tapering, particularly for patients who may require additional support due to limited health literacy or social isolation, such as patients living alone (Appleby and Al Musaimi 2025; Che et al. 2018; Risbo et al. 2025). Structured physiotherapy after discharge could aid recovery by ensuring safe rehabilitation and reducing reliance on opioids (Binder et al. 2004), especially for patients with higher pre-fracture functional status who might be more likely to overexert themselves (Tudorache et al. 2025). Engaging in early mobilisation during admission may improve recovery and lower the risk of LTOT (Agarwal et al. 2024; Tudorache et al. 2025). Lastly, involving the patient's support network in pain management planning, especially for those at elevated risk, may enhance adherence to tapering strategies and address psychosocial factors influencing opioid use (Che et al. 2018; Risbo et al. 2025).

Although the model is not perfect, it represents a first step toward developing a screening tool for LTOT risk in a frail, high-risk hip fracture population, where prolonged opioid use is a common and clinically relevant problem. Importantly, the model's development followed recommended prediction modelling guidelines, including variable selection, internal validation and reporting of discrimination, calibration and decision curve analysis metrics (Heinze et al. 2018; Kerr et al. 2016; Steyerberg and Vergouwe 2014), providing a robust foundation to justify external validation and feasibility testing in future studies prior to clinical implementation. These steps are essential to establish the model's practical utility and effectiveness in clinical decision-making.

5 | Conclusion

This study presents a novel prediction model for postoperative LTOT specifically in opioid-naïve hip fracture patients. By identifying key risk factors and demonstrating potential clinical utility, the model provides a first step toward a tool for risk stratification at discharge. While promising, the model is not yet ready for routine clinical use, and external validation and feasibility studies are needed to assess the model's generalisability and impact on reducing patient rates of LTOT.

Author Contributions

Conceptualisation: A.B.P. and Y.M.T. Statistical analyses: Y.M.T. Project administration: A.B.P. and Y.M.T. Supervision and input to statistical analyses: All authors. Writing – original draft: Y.M.T. Writing – review and editing: All authors.

Funding

This work was supported by the Department of Clinical Medicine, Aarhus University.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

To protect the privacy of patients, it is by Danish Data Law prohibited to make individual level data public. Access may be granted only by relevant Danish authorities.

References

Agarwal, N., T. Feng, A. Maclullich, A. Duckworth, and N. Clement. 2024. "Early Mobilisation After Hip Fracture Surgery Is Associated With Improved Patient Outcomes: A Systematic Review and Meta-Analysis." *Musculoskeletal Care* 22, no. 1: e1863. https://doi.org/10.1002/msc.1863.

Appleby, S., and O. Al Musaimi. 2025. "Socioeconomic Deprivation and Opioid Consumption: An Analysis Across England." *International Journal of Environmental Research and Public Health* 22, no. 5: 750. https://doi.org/10.3390/ijerph22050750.

Benyamin, R., A. M. Trescot, S. Datta, et al. 2008. "Opioid Complications and Side Effects." *Pain Physician* 11, no. 2 Suppl: S105–S120. https://www.ncbi.nlm.nih.gov/pubmed/18443635.

Bhandari, M., and M. Swiontkowski. 2017. "Management of Acute Hip Fracture." *New England Journal of Medicine* 377, no. 21: 2053–2062. https://doi.org/10.1056/NEJMcp1611090.

Binder, E. F., M. Brown, D. R. Sinacore, K. Steger-May, K. E. Yarasheski, and K. B. Schechtman. 2004. "Effects of Extended Outpatient Rehabilitation After Hip Fracture A Randomized Controlled Trial." *JAMA* 292, no. 7: 837–846. https://doi.org/10.1001/jama.292.7.837.

Chaudhary, M. A., N. Bhulani, E. C. de Jager, et al. 2019. "Development and Validation of a Bedside Risk Assessment for Sustained Prescription Opioid Use After Surgery." *JAMA Network Open* 2, no. 7: e196673. https://doi.org/10.1001/jamanetworkopen.2019.6673.

Che, X., R. Cash, S. K. Ng, P. Fitzgerald, and B. M. Fitzgibbon. 2018. "A Systematic Review of the Processes Underlying the Main and the Buffering Effect of Social Support on the Experience of Pain." *Clinical Journal of Pain* 34, no. 11: 1061–1076. https://doi.org/10.1097/ajp.0000000000000624.

Chou, R., D. B. Gordon, O. A. de Leon-Casasola, et al. 2016. "Management of Postoperative Pain: A Clinical Practice Guideline From the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative

Council." *Journal of Pain* 17, no. 2: 131–157. https://doi.org/10.1016/j.jpain.2015.12.008.

Christodoulou, E., J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster. 2019. "A Systematic Review Shows no Performance Benefit of Machine Learning Over Logistic Regression for Clinical Prediction Models." *Journal of Clinical Epidemiology* 110: 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004.

Collins, G. S., K. G. M. Moons, P. Dhiman, et al. 2024. "TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods." *BMJ* 385: e078378. https://doi.org/10.1136/bmj-2023-078378.

Cupp, M. A., F. L. Beaudoin, K. N. Hayes, et al. 2023. "Post-Acute Care Setting After Hip Fracture Hospitalization and Subsequent Opioid Use in Older Adults." *Journal of the American Medical Directors Association* 24, no. 7: 971–977.e974. https://doi.org/10.1016/j.jamda.2023.03.012.

Edwards, N. M., C. Varnum, S. Overgaard, L. Nikolajsen, C. F. Christiansen, and A. B. Pedersen. 2021. "Risk Factors for New Chronic Opioid Use After Hip Fracture Surgery: A Danish Nationwide Cohort Study From 2005 to 2016 Using the Danish Multidisciplinary Hip Fracture Registry." *BMJ Open* 11, no. 3: e039238. https://doi.org/10.1136/bmjopen-2020-039238.

Gadgaard, N. R., C. Varnum, R. Nelissen, C. Vandenbroucke-Grauls, H. T. Sørensen, and A. B. Pedersen. 2024. "Major Comorbid Diseases as Predictors of Infection in the First Month After Hip Fracture Surgery: A Population-Based Cohort Study in 92,239 Patients." *Eur Geriatr Med* 15, no. 4: 1069–1080. https://doi.org/10.1007/s41999-024-00989-w.

Heinze, G., C. Wallisch, and D. Dunkler. 2018. "Variable Selection – A Review and Recommendations for the Practicing Statistician." *Biometrical Journal* 60, no. 3: 431–449. https://doi.org/10.1002/bimj.201700067.

Hereford, T. E., A. Porter 3rd, J. B. Stambough, S. M. Cherney, and S. C. Mears. 2022. "Prevalence of Chronic Opioid Use in the Elderly After Hip Fracture Surgery." *Journal of Arthroplasty* 37, no. 7S: S530–S535. https://doi.org/10.1016/j.arth.2022.01.071.

Hjelholt, T. J., S. P. Johnsen, P. K. Brynningsen, J. S. Knudsen, D. Prieto-Alhambra, and A. B. Pedersen. 2022. "Development and Validation of a Model for Predicting Mortality in Patients With Hip Fracture." *Age and Ageing* 51, no. 1: afab233. https://doi.org/10.1093/ageing/afab233.

Johnson, A., B. Milne, N. Jamali, et al. 2022. "Chronic Opioid Use After Joint Replacement Surgery in Seniors Is Associated With Increased Healthcare Utilization and Costs: A Historical Cohort Study." *Canadian Journal of Anaesthesia* 69, no. 8: 963–973. https://doi.org/10.1007/s12630-022-02240-1.

Karhade, A. V., T. D. Cha, H. A. Fogel, et al. 2020. "Predicting Prolonged Opioid Prescriptions in Opioid-Naïve Lumbar Spine Surgery Patients." *Spine Journal* 20, no. 6: 888–895. https://doi.org/10.1016/j.spinee.2019.12.019.

Karhade, A. V., J. H. Schwab, and H. S. Bedair. 2019. "Development of Machine Learning Algorithms for Prediction of Sustained Postoperative Opioid Prescriptions After Total Hip Arthroplasty." *Journal of Arthroplasty* 34, no. 10: 2272–2277.e2271. https://doi.org/10.1016/j.arth.2019.06.013.

Karmali, R. N., C. Bush, S. R. Raman, C. I. Campbell, A. C. Skinner, and A. W. Roberts. 2020. "Long-Term Opioid Therapy Definitions and Predictors: A Systematic Review." *Pharmacoepidemiology and Drug Safety* 29, no. 3: 252–269. https://doi.org/10.1002/pds.4929.

Katakam, A., A. V. Karhade, J. H. Schwab, A. F. Chen, and H. S. Bedair. 2020. "Development and Validation of Machine Learning Algorithms for Postoperative Opioid Prescriptions After TKA." *Journal of Orthopaedics* 22: 95–99. https://doi.org/10.1016/j.jor.2020.03.052.

Kehlet, H. 2020. "Enhanced Postoperative Recovery: Good From Afar, but Far From Good?" *Anaesthesia* 75, no. Suppl 1: e54–e61. https://doi.org/10.1111/anae.14860.

Kerr, K. F., M. D. Brown, K. Zhu, and H. Janes. 2016. "Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use." *Journal of Clinical Oncology* 34, no. 21: 2534–2540. https://doi.org/10.1200/jco.2015.65.5654.

Kristensen, P. K., N. D. Rock, H. C. Christensen, and A. B. Pedersen. 2020. "The Danish Multidisciplinary Hip Fracture Registry 13-Year Results From a Population-Based Cohort of Hip Fracture Patients." *Clinical Epidemiology* 12: 9–21. https://doi.org/10.2147/CLEP.S231578.

Lindestrand, A. G., M. L. Christiansen, C. Jantzen, S. van der Mark, and S. E. Andersen. 2015. "Opioids in Hip Fracture Patients: An Analysis of Mortality and Post Hospital Opioid Use." *Injury* 46, no. 7: 1341–1345. https://doi.org/10.1016/j.injury.2015.04.016.

Lynam, A. L., J. M. Dennis, K. R. Owen, et al. 2020. "Logistic Regression Has Similar Performance to Optimised Machine Learning Algorithms in a Clinical Setting: Application to the Discrimination Between Type 1 and Type 2 Diabetes in Young Adults." *Diagnostic and Prognostic Research* 4, no. 1: 6. https://doi.org/10.1186/s41512-020-00075-2.

Matos, A., D. L. Bankes, K. T. Bain, T. Ballinghoff, and J. Turgeon. 2020. "Opioids, Polypharmacy, and Drug Interactions: A Technological Paradigm Shift Is Needed to Ameliorate the Ongoing Opioid Epidemic." *Pharmacy (Basel)* 8, no. 3: 154. https://doi.org/10.3390/pharmacy8030154.

Mazarello Paes, V., A. Ting, J. Masters, M. V. I. Paes, S. M. Graham, and M. L. Costa. 2025. "A Systematic Review of the Association Between Early Comprehensive Geriatric Assessment and Outcomes in Hip Fracture Care for Older People." *Bone & Joint Journal* 107-B: 595–603. https://doi.org/10.1302/0301-620x.107b6.Bjj-2024-1255.R1.

Pedersen, A. B., E. M. Mikkelsen, D. Cronin-Fenton, et al. 2017. "Missing Data and Multiple Imputation in Clinical Epidemiological Research." *Clinical Epidemiology* 9: 157–166. https://doi.org/10.2147/clep.S129785.

Quaye, A., J. DiPalazzo, K. Kostka, et al. 2025. "Identifying Factors Associated With Persistent Opioid Use After Total Joint Arthroplasty: A Retrospective Review." *Pain Medicine* 26, no. 2: 53–62. https://doi.org/10.1093/pm/pnae120.

Reps, J. M., P. Rijnbeek, A. Cuthbert, P. B. Ryan, N. Pratt, and M. Schuemie. 2021. "An Empirical Analysis of Dealing With Patients Who Are Lost to Follow-Up When Developing Prognostic Models Using a Cohort Design." *BMC Medical Informatics and Decision Making* 21, no. 1: 43. https://doi.org/10.1186/s12911-021-01408-x.

Risbo, N., V. Ehrenstein, P. H. Gundtoft, J.-E. Gjertsen, and A. B. Pedersen. 2025. "Socioeconomic Position and Chronic Opioid Use After Hip Fracture Surgery: A Danish Population-Based Cohort Study." *European Journal of Pain* 29, no. 6: e70063. https://doi.org/10.1002/ejp.70063.

Schmidt, M., L. Pedersen, and H. T. Sørensen. 2014. "The Danish Civil Registration System as a Tool in Epidemiology." *European Journal of Epidemiology* 29, no. 8: 541–549. https://doi.org/10.1007/s10654-014-9930-3.

Schneeweiss, S., and J. Avorn. 2005. "A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics." *Journal of Clinical Epidemiology* 58, no. 4: 323–337. https://doi.org/10.1016/j.jclinepi.2004.10.012.

Simoni, A. H., L. Nikolajsen, A. E. Olesen, C. F. Christiansen, and A. B. Pedersen. 2019. "Opioid Use After Hip Fracture Surgery: A Danish Nationwide Cohort Study From 2005 to 2015." *European Journal of Pain* 23, no. 7: 1309–1317. https://doi.org/10.1002/ejp.1392.

Sing, C. W., T. C. Lin, S. Bartholomew, et al. 2023. "Global Epidemiology of Hip Fractures: Secular Trends in Incidence Rate, Post-Fracture Treatment, and All-Cause Mortality." *Journal of Bone and Mineral Research* 38: 1064–1075. https://doi.org/10.1002/jbmr.4821.

Steyerberg, E. W. 2019. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* Springer International Publishing. https://books.google.dk/books?id=d2WCwgEACAAJ.

Steyerberg, E. W., and Y. Vergouwe. 2014. "Towards Better Clinical Prediction Models: Seven Steps for Development and an ABCD for Validation." *European Heart Journal* 35, no. 29: 1925–1931. https://doi.org/10.1093/eurheartj/ehu207.

Stone, J. M., A. Pujari, J. Garlich, and C. Lin. 2023. "A Retrospective Cohort Study on Chronic Opioid Use After Geriatric Hip Fracture Surgery-Risk Factors, Trends, and Outcomes." *Journal of the American Academy of Orthopaedic Surgeons* 31, no. 6: 312–318. https://doi.org/10.5435/JAAOS-D-22-00458.

Sun, Z., W. Dong, H. Shi, H. Ma, L. Cheng, and Z. Huang. 2022. "Comparing Machine Learning Models and Statistical Models for Predicting Heart Failure Events: A Systematic Review and Meta-Analysis." *Frontiers in Cardiovascular Medicine* 9: 812276. https://doi.org/10.3389/fcvm.2022.812276.

Tudorache, Y. M., I. T. Andersen, T. J. Hjelholt, M. T. Kristensen, K. J. Sheehan, and A. B. Pedersen. 2025. "Association Between Early Mobilization After Hip Fracture Surgery and Risk of Long-Term Opioid Therapy." *European Geriatric Medicine* 16: 921–931. https://doi.org/10.1007/s41999-025-01227-7.

van Diepen, M., C. L. Ramspek, K. J. Jager, C. Zoccali, and F. W. Dekker. 2017. "Prediction Versus Aetiology: Common Pitfalls and How to Avoid Them." *Nephrology, Dialysis, Transplantation* 32, no. suppl_2: ii1–ii5. https://doi.org/10.1093/ndt/gfw459.

Vickers, A. J., and E. B. Elkin. 2006. "Decision Curve Analysis: A Novel Method for Evaluating Prediction Models." *Medical Decision Making* 26, no. 6: 565–574. https://doi.org/10.1177/0272989x06295361.

Yue, Q., Y. Ma, Y. Teng, et al. 2020. "An Updated Analysis of Opioids Increasing the Risk of Fractures." *PLoS One* 15, no. 4: e0220216. https://doi.org/10.1371/journal.pone.0220216.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** ejp70280-sup-0001-Supinfo01.xlsx. **Figure S1:** Schneeweiss-inspired study design diagram. **Table S1:** Coding scheme for the opioids included in the outcome definition. **Table S2:** Coding scheme for the hip fracture and operation definitions. **Table S3:** Coding scheme for preoperative medication. **Table S4:** Coding scheme for preoperative comorbidities. **Table S5:** Machine learning algorithms, corresponding R packages and their settings. **Figure S2:** Predictor distribution among patients with and without long-term opioid therapy. **Figure S3:** Receiver operating characteristic curves for the machine learning algorithms used in this study. **Table S6:** Baseline patient characteristics comparison between complete case patients used in the study and patients with missing predictor data.