Development and internal validation of a prediction model for long-term opioid use-an analysis of insurance claims data

Held, Ulrike ; Forzy, Tom ; Signorell, Andri ; Deforth, Manja ; Burgstaller, Jakob M ; Wertli, Maria M

DOI: https://doi.org/10.1097/j.pain.0000000000003023

Research Paper

PAIN 165 (2024) 44–53

PAIN®

VIDEO EC C OPEN

Development and internal validation of a prediction model for long-term opioid use—an analysis of insurance claims data

Ulrike Held[a,*], Tom Forzy[b], Andri Signorell[c], Manja Deforth[a], Jakob M. Burgstaller[d], Maria M. Wertli[e,f]

Abstract

In the United States, a public-health crisis of opioid overuse has been observed, and in Europe, prescriptions of opioids are strongly increasing over time. The objective was to develop and validate a multivariable prognostic model to be used at the beginning of an opioid prescription episode, aiming to identify individual patients at high risk for long-term opioid use based on routinely collected data. Predictors including demographics, comorbid diseases, comedication, morphine dose at episode initiation, and prescription practice were collected. The primary outcome was long-term opioid use, defined as opioid use of either >90 days duration and ≥10 claims or >120 days, independent of the number of claims. Traditional generalized linear statistical regression models and machine learning approaches were applied. The area under the curve, calibration plots, and the scaled Brier score assessed model performance. More than four hundred thousand opioid episodes were included. The final risk prediction model had an area under the curve of 0.927 (95% confidence interval 0.924-0.931) in the validation set, and this model had a scaled Brier score of 48.5%. Using a threshold of 10% predicted probability to identify patients at high risk, the overall accuracy of this risk prediction model was 81.6% (95% confidence interval 81.2% to 82.0%). Our study demonstrated that long-term opioid use can be predicted at the initiation of an opioid prescription episode, with satisfactory accuracy using data routinely collected at a large health insurance company. Traditional statistical methods resulted in higher discriminative ability and similarly good calibration as compared with machine learning approaches.

Keywords: Long-term opioid use, Morphine equivalent, Clinical prediction model, Validation, Insurance claims data, Pain medication, Chronic pain

1. Introduction

1.1. Background and objectives

In the United States, a public health crisis of opioid abuse and addiction as a result of liberal opioid regulation and use has been observed.[16] In Europe, the monitoring system indicates an increased use of opioids and there is some evidence of increased opioid addiction.[33] Increasing prescriptions of opioids in the United States were a result of physicians' accepting low-quality studies' evidence that opioids are effective and harmless to treat chronic pain.[3,20,32] Given the potential personal, societal, public health, and economic costs of opioid dependency and opioid-related adverse events, preventive measures should be implemented to prevent unintended long-term opioid use.

Opioids are intended to relieve acute severe pain in patients with active cancer, but strong opioids are not recommended for long-term use in patients with chronic noncancer pain since nonopioid alternatives exist.[9] In chronic pain, opioids are not more effective than nonopioid medications but may also have unintended adverse events.[23] Common adverse events after opioid prescriptions are emergency department visits, infections, hospitalizations, ICU admissions, or death, and these are known to occur with an increased risk with increase in daily dosage and with increased duration of opioid use.[1,3,36,40] Thus, although the short-term use of opioid therapy may be beneficial, long-term use of opioids is generally not associated with benefit and is associated with risk for these adverse events. Therefore, it would be clinically important to know at the initiation of an opioid prescription episode, who will be at high risk to evolve into long-term opioid use.

In this study, we aimed to develop and validate a prognostic multivariable clinical prediction model for the outcome long-term opioid use that may help to identify individual patients at risk for long-term opioid use. We used a patient cohort with opioid use and assessed a large set of risk factors for long-term opioid use.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] Department of Biostatistics at Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland, [b] Master Program Statistics, ETH Zurich, Zurich, Switzerland, [c] Department of Health Sciences, Helsana, Dübendorf, Switzerland, [d] Institute of Primary Care, University and University Hospital Zurich, Zurich, Switzerland, [e] Department of Internal Medicine, Cantonal Hospital Baden KSB, Baden, Switzerland, [f] Department of General Internal Medicine University Hospital Bern, University of Bern, Switzerland, Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

**Corresponding Author. Address. Department of Biostatistics at Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland. E-mail address: Ulrike.Held@uzh.ch (Ulrike Held).*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

http://dx.doi.org/10.1097/j.pain.0000000000003023

We hypothesized that socioeconomic, episode-specific, and prescription-specific factors, as well as comorbidities and comedication use may help to identify individuals at high risk for long-term opioid use.

2. Methods

The reporting of this study was performed according to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines.[6]

2.1. Source of data

This study is based on data of one of the largest insurance companies in Switzerland, the Helsana insurance group. The patient-level linked database provided information on sociodemographic data, prescribed drugs, and healthcare encounters.

2.2. Participants

The study cohort consisted of all consecutive adult patients (aged 18 years and older) of the Helsana insurance company with at least one opioid claim between January 1, 2013, and December 31, 2018. Opioid prescriptions were identified using unique codes of the WHO pharmacological Anatomical Therapeutic Chemical (ATC) classification system. Each opioid prescription is associated with a morphine conversion factor that can be used to calculate morphine equivalent doses to compare different opioids. Excluded were those patients who were in an opioid use disorder treatment program.[3]

2.3. Definition of an opioid episode

The methods used to develop opioid episodes have been previously described.[3] In brief, an opioid episode was defined as a continuous time interval in which we assumed daily opioid use. The first day of an episode begins with an initial opioid prescription, that is, claim. The duration of an opioid episode was calculated as the difference in days between the date of the initial prescription and the run-out date of the last prescription plus 1. If 3 months after the calculated run-out date no new claim was filed, the episode ended.

2.4. Average daily opioid dose at episode initiation

The average daily dose was calculated based on the morphine equivalent dose (MED) per prescription.[3,38] In cases of multiple claims, the MED per treatment day was calculated across the multiple claims. For the calculation of the initial dose category, the daily MED dose between claims was categorized into 4 groups of average daily dosage, <20, 20 to <50, 50 to <100, and ≥ 100 mg MED per day.

2.5. Outcome

The primary outcome to be predicted in this study was the development of long-term opioid use. Following the definition by von Korff et al.,[37] long-term opioid use is defined as an episode of opioid use of either >90 days duration and ≥ 10 opioid claims or >120 days, independent of the number of opioid claims. All other episodes of opioid prescriptions were considered short-term (ie, acute or subacute). The outcome categorization was performed retrospectively for all patients included in the analysis. It was assessed independently from the predictor variables at episode start.

2.6. Predictors

The predictors considered in this study covered different aspects of risk for long-term opioid use in each patient. Predictors were defined a priori by consulting the relevant literature[17] and during discussions within the interdisciplinary research group. There was no variable selection performed based on methodological criteria. The timing when each of the predictors was assessed was at the beginning of each opioid episode since the prediction model is intended to be used at the time of episode initiation.

2.7. Predictor domains included

(1) Demographic variables: age in categories <50 years (reference category), 50 to <60, 60 to <70, 70 to <80, and ≥ 80 and sex
(2) Socioeconomic variables: place of residence in a non-German language speaking vs German-speaking canton of Switzerland, insurance type (additional private or semiprivate insured), and managed care model
(3) Episode specific variables: initial dosage category categorized into <20, 20 to <50, 50 to <100, and ≥ 100 mg MED per day; prescriber variables (indicating whether the patient had a single vs multiple prescribers of opioids); previous opioid use categorized into never, >2 years ago, 6 months to 2 years ago, and within last 6 months
(4) Disease specific risk factors: comorbidities and comedication use.

Although in Switzerland, the level of inequality is relatively low compared with other countries,[10] there are some differences in social status. The socioeconomic variables include proxies such as the place of residence (French/Italian-speaking parts as compared with German-speaking parts of Switzerland are associated with a lower index of socioeconomic position),[26] additional private or semiprivate insurance is more expensive, and managed care models are less expensive than standard policies.

Comorbid diseases were identified using the Chronic Disease Score (CDS). The CDS has been associated with healthcare use.[15,19,28] Chronic diseases included chronic infections, inflammatory disease, renal disease, endocrinologic disease, diabetes, lung or pulmonary disease, neurological disease, cardiovascular disease, hyperlipidemia, glaucoma, acid peptic disease, thyroid disease, gout, psychiatric disease or depression, spine disease, musculoskeletal disease, and cancer. Comorbid diseases were included as binary predictors.

Comedication use included stimulants, bisphosphonates, muscle relaxants, nonopioid analgesic use, and benzodiazepines, all of them also coded as binary predictors assessed at episode initiation.

2.8. Sample size

The initial number of patients for the development and validation of the prediction model was 266,476. This allows the simultaneous evaluation of all the predictor variables without any restrictions.[30]

2.9. Missing data

The number of patients with missing values in one or more of the predictor variables was 53 patients, representing less than 0.02% of the patients. Given the computational complexity of a methodological solution for missing values in a database of this size, the decision to exclude these patients seemed adequate.

2.10. Statistical analysis methods

All analyses were performed with the use of the statistical programming language R[29] (version 4.2.0) and specific additional packages, in a fully scripted and reproducible way. In the primary analysis, a multiple logistic regression model (generalized linear model) with fixed effects was fitted to the outcome of long-term opioid use. In this model, no interaction terms between predictor variables were evaluated. No variable selection was performed. Model assumptions were verified.

As a sensitivity analysis, a generalized linear mixed effects model with random intercepts was fitted to account for the correlation of repeated episodes within patients. We also performed another sensitivity analysis using machine learning approaches. We fitted 2 random forest models using the same predictor variables and outcome as before. Random forest models are bagged decision tree models that are fitted on subsets of the full data set. The use of random forest models can be justified by their flexibility. Random forest models are typically adequate for large data sets, are robust to outliers, and still allow for the assessment of relative variable importance. The first random forest model was fitted on default bootstrap samples, meaning that the training sets for the decision trees were sampled randomly with replacement from the full data set. To control imbalance in the initial data set regarding the outcome categories, a second random forest model was fitted on stratified random samples. This means that the data set is divided in strata before the random sampling. In the random forest framework, potential interactions are naturally considered. For fitting the random forests, the R-package randomForest[39] was used.

2.11. Model derivation and validation, model performance

We used an internal–external validation approach for the logistic regression model, and the random forests, by repeatedly splitting the data set into a training and a validation set of sizes 90% and 10%, respectively. Model performance was assessed for discrimination ability and calibration across validation sets. Discrimination was quantified with the area under the receiver operating characteristic curve (AUC) with bootstrap 95% confidence intervals (CI).[4] Calibration was visualized with calibration plots by dividing the obtained predictions of event for every individual into groups and then measuring how close the average prediction is to the actual proportion of events within the group. The calibration plots show the observed event rate as a function of the predicted rate. We have one point for every group, from which we create a smooth calibration curve by fitting restricted cubic splines using 3 knots. Calibration was assessed overall and in subgroups of patients with different covariate patterns, as suggested by Van Calster et al.[35] The scaled Brier score was calculated as a measure to quantify overall accuracy of the predicted probabilities. The scaled Brier score typically ranges from 0% to 100%, and its interpretation is that higher values indicate better overall accuracy. The scaled Brier score's interpretation is similar to Pearson R^2 statistic, and thus, resulting estimates can be interpreted accordingly.[13] In addition, mean decrease in accuracy and mean decrease in Gini coefficient were evaluated graphically for the random forest approach with random sampling.

2.12. Risk groups and thresholding

A threshold of 10% predicted probability for long-term opioid use in the final logistic regression model was chosen to classify episodes into high risk vs low risk. The threshold was chosen because a risk for long-term opioid use of 10% may be considered acceptable when assuming that a conservative estimate of 2.5% of those patients will eventually develop an opioid use disorder.[18] The thresholding may simplify a clinical decision based on the risk prediction model's probabilities estimated from logistic regression. Patients with a predicted probability of ≥10% were considered at high risk for the development of a long-term opioid episode, whereas patients below the threshold were considered at low risk. This allowed the direct comparison between the results from the logistic regression analysis with the random forest approaches. Overall accuracy, as measured with percentage of correctly classified episodes, as well as sensitivity, specificity, and positive and negative predictive value were estimated. The same threshold of ≥ 10% of the derivation set was also used in the validation set.

3. Results

Between January 1, 2013, and December 31, 2018, 418,625 episodes of opioid prescriptions were observed in a population of 266,476 patients. Of these, 61 episodes (53 patients) had missing data for one or more predictor variables and were excluded from the analyses (**Fig. 1**). In total, 418,564 (71,863 [17%] long-term) episodes in 266,423 patients were analyzed. The median duration of an episode was 7 days (interquartile range, IQR, 3-50 days), and the mean duration was 106 days (standard deviation, SD = 301 days).

Patient characteristics of the episodes at the initiation of opioid use are summarized in **Table 1**. Most patients were 70 years or older (37.1%), female (59.7%), living in a German-speaking region (71.4%), and had no additional insurance or managed care model. Most episodes started with opioid doses <20 mg MED (39.4%) and only in a minority of episodes multiple prescribers were observed (1.6%). Most patients had no prior episode (63.7%). Most frequently observed comorbid diseases in the overall population were inflammatory diseases (20.3%), cardiovascular disease (29.1%), acid peptic disease (32.3%), and musculoskeletal disease (13.6%). Comedications included nonopioid analgesics (66.0%), muscle relaxants (11.1%), and benzodiazepines (12.5%).

A higher proportion of patients with long-term episodes were aged 70 years or older (51.5% vs 34.1%), had episodes with an initial opioid dose <20 mg MED (56.0% vs 36.0%), and had a prior episode of less than 6 months ago (20.6% vs 9.6%). Patients with long-term episodes had more inflammatory diseases (49.1% vs 14.3%), lung diseases (19.5% vs 3.2%), diabetes (16.2% vs 3.6%), cardiovascular disease (72.6% vs 20.0%), acid peptic disease (71.4% vs 24.2%), psychiatric diseases or depression (43.6% vs 6.1%), musculoskeletal disease (40.5% vs 8.0%), and cancer (13.1% vs 6.4%). Patients with long-term episodes were more likely to have additional nonopioid analgesics (90.7% vs 60.8%) and benzodiazepines (36.5% vs 7.5%). Patients with short-term opioid episodes were younger (48.9% aged 59 or younger), in a managed care model (51.0% vs 40.3%), had a higher proportion with initial MED dose between 50 and 100 mg (22.4% vs 12.3%), and had no prior episode (65.2% vs 56.1%).

3.1. Logistic regression model

The final logistic regression model with fixed effects (**Table 2**) showed an increased risk for long-term opioid use in patients aged 80 years and older (OR 1.525, 95% CI 1.468-1.585), with multiple prescribers (OR 1.222, 95% CI 1.124-1.328), and a previous episode within the last 6 months (OR 1.789, 95% CI

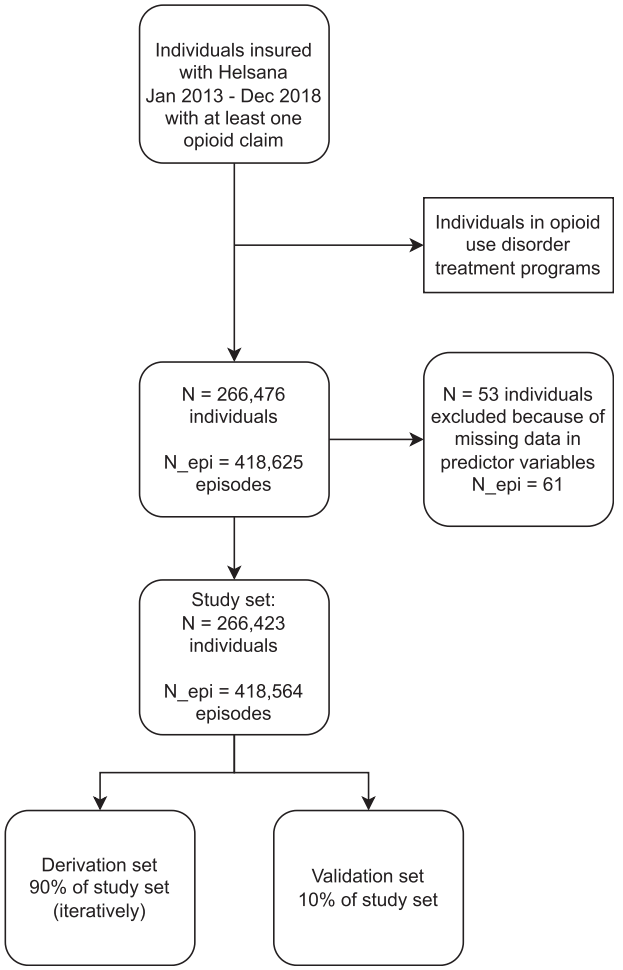

Figure 1. Flow chart of study population and study set used for analysis. N_epi indicates the number of episodes. N indicates the number of patients.

1.732-1.848). The presence of comorbid diseases and comedications increased the odds for long-term opioid use.

The final fixed-effects model resulted in an AUC value of 0.927 (95% CI 0.924-0.931), **Figure 2**, and a scaled Brier score of 48.5% in the validation set. As a sensitivity analysis, random forests with bootstrap samples and with stratified samples were trained. The resulting AUC values were 0.909 (0.905-0.913) and 0.921 (0.918-0.925), respectively. Mean decrease in accuracy (eFig. 1, available at http://links.lww.com/PAIN/B902) and mean decrease in Gini coefficient (eFig. 2, available at http://links.lww.com/PAIN/B902) were evaluated graphically for the random forest with bootstrap samples. Both plots indicate that the predictors' psychiatric disease or depression and cardiovascular disease are most relevant for accuracy and homogeneity of the random forest, respectively. The ROC curves of the random forest approaches can be found in eFig. 3 (available at http://links.lww.com/PAIN/B902). The scaled Brier scores were 46.0% and 29.1%, respectively.

Regarding the models' calibration, the calibration plots for the fixed effects logistic regression, the random forest with bootstrap samples, and the random forest with stratified sample were shown and revealed that the calibration of the former 2 models (**Fig. 3** left and middle) were comparably good, whereas the latter calibration plot (**Fig. 3** right) showed unsatisfactory calibration. Calibration was assessed in subgroups of patients with different covariate patterns. In eFig. 4 (available at http://links.lww.com/PAIN/B902), calibration for the fixed-effects logistic regression model is shown in subgroups of sex and age groups.

3.2. Thresholding

Using a threshold of ≥ 10% predicted probability of long-term opioid use, indicating patients at high risk, the overall accuracy of the fixed-effects logistic regression was 81.6% (95% CI from 81.2% to 82.0%). The corresponding sensitivity was 90.5%, and specificity was 80.8%, resulting in a positive predictive value of 48.5% and negative predictive value of 97.5%. The corresponding results for the random forest with bootstrap samples were 86.5% overall accuracy (95% CI from 86.1 to 86.8), sensitivity of 83.2%, and specificity of 87.2%. For the selected random forest with stratified samples, the corresponding overall accuracy was 77.5% (95% CI from 77.1 to 77.9), sensitivity was 93.4%, and specificity was 74.2%.

3.3. How to use the prediction model

The prediction model resulting from the logistic regression analysis can be used by multiplying the risk factor information of each new patient in the initial phase of an opioid episode with the corresponding coefficient as shown in **Table 2** (including the intercept) and adding them up, resulting in S. The result needs to be transformed to the probability scale by calculating $y = \frac{\exp(S)}{1+\exp(S)}$ as the returned individual's predicted probability for the episode developing into one with long-term opioid use.

In this context, 6 patient scenarios were derived to demonstrate the usefulness of the model for clinical practice. A low-risk scenario would be a 65-year-old male person, living in the Italian-speaking part of Switzerland, is in a managed care model, has a semiprivate insurance, received an initial dose of 30 mg from one prescriber, had no prior episode, does not have any comorbid diseases, and takes no comedication. For this man, the predicted probability to develop a long-term episode is extremely low (0.44%). A high-risk scenario would be a 70-year-old woman, living in the German-speaking part of Switzerland, is in no managed care model, with an initial dose of 30 mg from one prescriber, with a prior episode of opioid use within the last 6 months, with lung disease, musculoskeletal and neurological disease, and currently taking bisphosphonates and benzodiazepines at the start of the opioid episode. For this woman, the predicted probability to develop a long-term opioid episode is with 62% very high. More scenarios are presented in the style of a study by Oliva et al.,[25] together with the resulting predicted probabilities for long-term opioid use, in **Table 3**.

4. Discussion

The prevalence of chronic pain in the U.S. adult population has been estimated to be 20%, with a dramatic increase in the elderly.[21,41] Older adults are also more likely to need surgery, a risk factor for initiating an opioid therapy. Owing to an impaired kidney function, other pain medications may not be an option, and thus, opioids are needed to control pain. Although opioids in chronic pain are no more effective than other pain medications, stopping opioid treatment has been shown to be challenging.[22] In our study, we showed that the development of long-term opioid use can be predicted with satisfactory accuracy, as measured with discriminative ability and calibration, if demographic information, episode specific information, comorbidities, and comedication use are known. In our model, comorbidities (chronic, inflammatory and lung diseases, diabetes,

Table 1

Descriptive statistics at episode initiation among long-term and short-term opioid users.

Variable	Long-term opioid use (N = 71,863)	Only short-term opioid use (N = 346,701)	Overall (N = 418,564)
Age group			
<50 years	11,709 (16.3%)	110,174 (31.8%)	121,883 (29.1%)
50 to <60 years	10,843 (15.1%)	59,131 (17.1%)	69,974 (16.7%)
60 to <70 years	12,294 (17.1%)	59,374 (17.1%)	71,668 (17.1%)
70 to <80 years	16,464 (22.9%)	61,634 (17.8%)	78,098 (18.7%)
≥80 years	20,553 (28.6%)	56,388 (16.3%)	76,941 (18.4%)
Sex			
Female	46,740 (65.0%)	202,983 (58.5%)	249,723 (59.7%)
Male	25,123 (35.0%)	143,718 (41.5%)	168,841 (40.3%)
Language region			
French or Italian	19,704 (27.4%)	99,933 (28.8%)	119,637 (28.6%)
Swiss German	52,159 (72.6%)	246,768 (71.2%)	298,927 (71.4%)
(Semi) private insurance	13,012 (18.1%)	64,393 (18.6%)	77,405 (18.5%)
Managed care model	28,972 (40.3%)	176,978 (51.0%)	205,950 (49.2%)
Initial MED mg dose category			
<20 (category D)	40,214 (56.0%)	124,859 (36.0%)	165,073 (39.4%)
20 to <50 (category C)	15,822 (22.0%)	124,078 (35.8%)	139,900 (33.4%)
50 to <100 (category B)	8863 (12.3%)	77,669 (22.4%)	86,532 (20.7%)
≥100 (category A)	6964 (9.7%)	20,095 (5.8%)	27,059 (6.5%)
Multiple prescribers	1376 (1.9%)	5473 (1.6%)	6849 (1.6%)
Prior episode			
No prior episode	40,310 (56.1%)	226,113 (65.2%)	266,423 (63.7%)
Previous episode <6 months	14,821 (20.6%)	33,403 (9.6%)	48,224 (11.5%)
Previous episode 6 months to 2 years	13,573 (18.9%)	63,193 (18.2%)	76,766 (18.3%)
Previous episode >2 years	3159 (4.4%)	23,992 (6.9%)	27,151 (6.5%)
Comorbid diseases			
Chronic infections	6329 (8.8%)	4243 (1.2%)	10,572 (2.5%)
Inflammatory disease	35,290 (49.1%)	49,713 (14.3%)	85,003 (20.3%)
Renal disease	943 (1.3%)	633 (0.2%)	1576 (0.4%)
Endocrinologic disease	1197 (1.7%)	776 (0.2%)	1973 (0.5%)
Diabetes	11,663 (16.2%)	12,446 (3.6%)	24,109 (5.8%)
Lung disease	14,048 (19.5%)	11,111 (3.2%)	25,159 (6.0%)
Neurological disease	4700 (6.5%)	2897 (0.8%)	7597 (1.8%)
Cardiovascular disease	52,184 (72.6%)	69,415 (20.0%)	121,599 (29.1%)
Hyperlipidemia	20,941 (29.1%)	16,929 (4.9%)	37,870 (9.0%)
Glaucoma	6108 (8.5%)	2749 (0.8%)	8857 (2.1%)
Acid peptic disease	51,278 (71.4%)	83,771 (24.2%)	135,049 (32.3%)
Thyroid disease	7938 (11.0%)	6430 (1.9%)	14,368 (3.4%)
Gout	4475 (6.2%)	3268 (0.9%)	7743 (1.8%)
Psychiatric dis. or depression	31,317 (43.6%)	21,129 (6.1%)	52,446 (12.5%)
Spine disease	7739 (10.8%)	3958 (1.1%)	11,697 (2.8%)
Musculoskeletal disease	29,114 (40.5%)	27,743 (8.0%)	56,857 (13.6%)
Cancer	9443 (13.1%)	22,315 (6.4%)	31,758 (7.6%)
Comedications			
Stimulants	637 (0.9%)	417 (0.1%)	1054 (0.3%)
Bisphosphonates	5473 (7.6%)	2457 (0.7%)	7930 (1.9%)
Muscle relaxants	13,711 (19.1%)	32,877 (9.5%)	46,588 (11.1%)
Nonopioid analgesics	65,194 (90.7%)	210,903 (60.8%)	276,097 (66.0%)
Benzodiazepines	26,243 (36.5%)	26,005 (7.5%)	52,248 (12.5%)

cardiovascular and acid peptic disease, musculoskeletal disease, psychiatric disease or depression, and cancer) were associated with an increased risk for long-term opioid use. Other factors that increased risk for long-term opioid use were patients aged 80 years and older, multiple prescribers, a previous episode within the past 6 months, and comedication use. Higher initial opioid dose was associated with reduced risk of long-term opioid use.

The results of our study are in line with similar models that were recently published.[2] Oliva et al.[25] derived a predictive risk model for the outcome overdose-related or suicide-related event in a Veterans Health Administration population with an opioid prescription. They found that mental health disorders including posttraumatic stress, major depression, bipolar disorder, and other mental health disorders increased the odds for overdose or suicide-related events substantially. In our study, psychiatric disease or depression were found to be the comorbidities with the largest estimated effect as compared with all other comorbidities. Sullivan and Howe[34] argue that

Table 2

Estimated coefficients, standard errors, odds ratios, and 95% confidence intervals of the final logistic regression model.

Domain	Reference	Label	Coefficient	SE	OR	95% CI for OR
		Intercept	−4.069	0.024	0.017	0.016–0.018
Demographic						
Age group	< 50 years	50 to <60 years	−0.031	0.019	0.969	0.933–1.007
		60 to <70 years	−0.151	0.02	0.860	0.827–0.894
		70 to <80 years	−0.099	0.02	0.906	0.872–0.942
		≥80 years	0.422	0.02	1.525	1.468–1.585
Sex	Female	Male	0.018	0.012	1.018	0.994–1.043
Socioeconomic						
Language region	Swiss German	French or Italian	−0.068	0.013	0.934	0.91–0.959
Insurance	No additional insurance	(Semi) private insurance	−0.193	0.016	0.824	0.799–0.85
Managed care	No managed care model	Managed care model	−0.096	0.012	0.909	0.887–0.93
Episode specific						
Initial MED (mg) dose category	<20	≥100	−0.567	0.022	0.567	0.543–0.593
		50 to <100	−1.276	0.017	0.279	0.27–0.289
		20 to <50	−0.869	0.014	0.420	0.408–0.432
Prescriber	One prescriber	>1 prescriber	0.2	0.043	1.222	1.124–1.328
Prior episode	No prior episode	Previous episode <6 months	0.582	0.017	1.789	1.732–1.848
		Previous episode 6 months to 2 years	−0.023	0.015	0.977	0.948–1.007
		Previous episode >2 years	−0.313	0.026	0.731	0.694–0.77
Disease specific						
Comorbidities	No comorbidities	Chronic infections	0.926	0.029	2.525	2.384–2.675
		Inflammatory disease	0.578	0.014	1.783	1.735–1.831
		Renal disease	0.196	0.079	1.217	1.043–1.42
		Endocrinologic disease	0.638	0.069	1.892	1.651–2.167
		Diabetes	0.453	0.02	1.573	1.513–1.636
		Lung disease	0.894	0.02	2.446	2.353–2.542
		Neurological disease	0.913	0.035	2.493	2.329–2.668
		Cardiovascular disease	1.165	0.013	3.206	3.122–3.291
		Hyperlipidemia	0.72	0.017	2.055	1.989–2.123
		Glaucoma	1.232	0.033	3.429	3.215–3.657
		Acid peptic disease	0.892	0.012	2.44	2.382–2.5
		Thyroid disease	0.665	0.024	1.944	1.853–2.04
		Gout	0.575	0.033	1.777	1.665–1.896
		Psychiatric disease or depression	1.381	0.014	3.981	3.871–4.093
		Spine disease	0.97	0.03	2.639	2.489–2.797
		Musculoskeletal disease	1.129	0.014	3.093	3.007–3.182
		Cancer	0.128	0.02	1.136	1.093–1.181
Comedications	No comedications	Stimulants	1.107	0.094	3.025	2.514–3.64
		Bisphosphonates	1.204	0.035	3.332	3.113–3.565
		Muscle relaxants	0.468	0.017	1.597	1.544–1.651
		Non-opioid analgesic use	0.86	0.017	2.363	2.287–2.441
		Benzodiazepines	0.802	0.015	2.23	2.167–2.295

CI, confidence interval; OR, odds ratio; SE, standard error.

long-term opioid use is recommended only for patients with persistent pain in the absence of a history of substance abuse. Although individuals in opioid use disorder treatment programs were not included in our study, we found that comedication use such as benzodiazepines and stimulants increased the risk for long-term opioid use. Apart from previous opioid use, we did not have any other documentation of individuals who might have a history of substance abuse in our study. Sullivan and Howe[34] describe a seemingly adverse selection of patients for long-term opioid use in which patients with substance use or mental health disorders are overrepresented in a population of opioid users.

Other publications were focused on the identification on strongest predictors among the set of potential predictor variables, and they found that older age, White race, hourly wage, low back pain, and osteoarthritis were the strongest predictors besides calendar year.[27] Another publication identified comorbidities to be indicative of incident chronic use, but the comparison group was patients without opioid prescriptions, therefore, not directly comparable with our study.[24] None of the above publications reported on overall performance of their models regarding discrimination or calibration, and their final models were not validated. In our application, logistic regression with a generalized linear model showed better discriminative ability than the 2 machine learning approaches. Calibration was comparable across the logistic regression model and the random forest with bootstrap samples. These findings are in line with systematic evaluations of Gravesteijn et al.[11] in which the authors found that the performance of machine learning approaches was generally not better than regression-based approaches in the field of traumatic brain injury. Christodoulou et al.[5] systematically evaluated the literature on clinical prediction models to evaluate whether there was evidence for superior performance of machine learning over logistic regression for binary outcomes. In clinical prediction models at

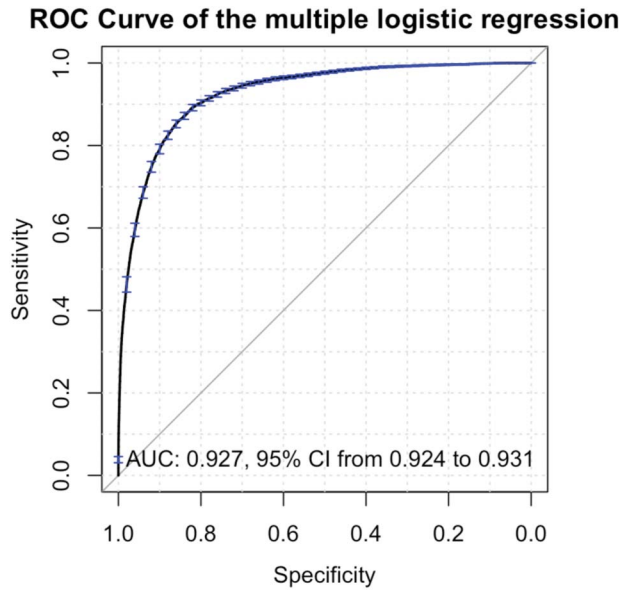

Figure 2. ROC curve of the logistic regression model (GLM).

low risk of bias, the authors found no evidence for a performance benefit of machine learning approaches. Calibration levels for risk prediction models were recently defined,[35] and our final model's calibration plots shows that predicted risks correspond well to observed event rates, not only overall but also for age groups and sex patterns, indicating a strong calibration level.

The population in our study included cancer and noncancer patients, whereas in other publications, the cancer patients were explicitly excluded. This makes our prediction model useful for all settings in which opioid use may be indicated. With a threshold of 10% or higher as cut-off for patients at high-risk for long-term opioid use, our model showed high accuracy overall, as well as high sensitivity and specificity.

4.1. Limitations and strengths

The data used for the derivation and validation of the prediction model were retrospective, and initially, the data were not collected for the purpose of the development of a clinical prediction model for long-term opioid use. The data were collected at Helsana health insurance group covering 1,2 million individuals and representing approximately 12 to 14% of the Swiss population across all 26 administrative regions (cantons) of Switzerland. A health insurance is compulsory in Switzerland. The data underlying the study may be seen as representative because comparisons of the Helsana population against the Swiss population revealed no evidence for differences in age and sex distribution. Helsana has good coverage across all Swiss cantons, and it was deemed representative for the Swiss population in recent publications on laboratory testing and antidepressant prescriptions.[12,14] In Switzerland, opioid prescriptions require a separate narcotic prescription that is filed by the pharmacy and reimbursed by the insurance company (tiers payant). Opioids are not sold over the counter, and thus, a high level of ascertainment is guaranteed.

Insurance claims data do not provide information on clinical diagnosis and the reasons for the opioid prescription. Although using insurance or routinely collected data often are not ideal and have limitations, they provide real-world evidence; they are readily available and may help to identify patients at risk for long-term opioid use. We assessed the burden of disease in patients by using CDC categories that are derived from individual medication use associated with specific comorbidities. The final prediction model had a high scaled Brier score of 48.5%, with an interpretation corresponding to that of an R^2 value of a linear regression model, and thus, it is rather unlikely that additional essential areas of information were missing. In our definition of an opioid episode, the actual opioid intake could be overestimated, if a patient did not use all prescribed opioids. In the assessment of whether a prior episode of opioid consumption was observed, there may be underreporting in the sense that patients might have had prior episodes in times where they were not covered by Helsana health insurance. We set a threshold for high-risk patients to develop long-term opioid use at 10% predicted probability. This threshold was based on an assumed estimate of 2.5% of patients eventually developing an opioid use disorder. Although a definition of a threshold for the overall risk for adverse events may be preferable, this may result in a very high number of patients at risk. According to a Cochrane systematic review, between 6 and 85% of patients suffered from any opioid related adverse event and between 2.5 and 22% from a serious adverse event.[8]

Our study has several strengths. These include the representativeness of the large number of observations collected across Switzerland. The number of events per variable was sufficient to fit

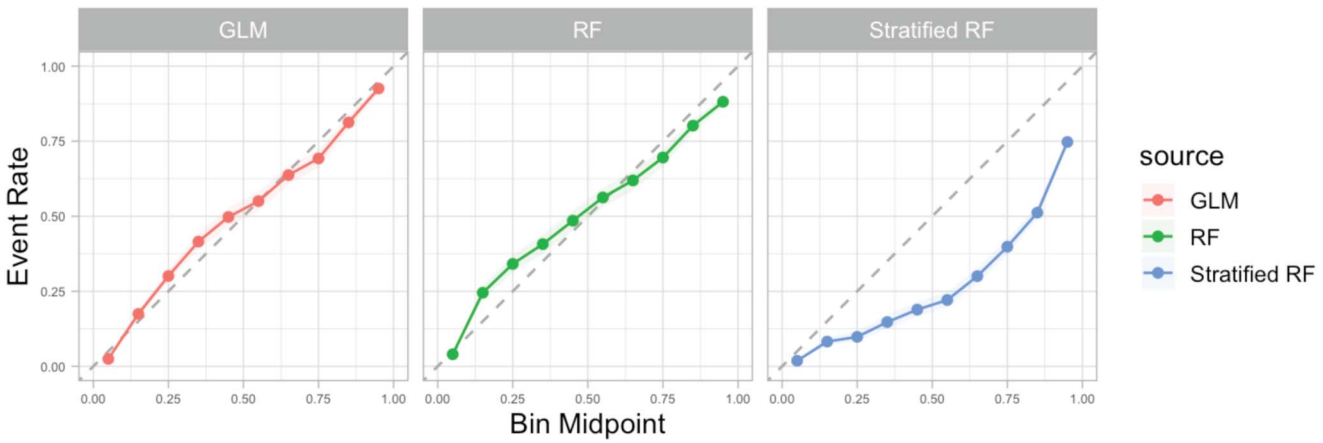

Figure 3. Faceted calibration plots, left: logistic regression (GLM), middle: random forest with bootstrap samples (RF), right: random forest with stratified samples (stratified RF). On the x-axis, bin midpoints represent the central values within each bin or interval; they are computed as the mean value of the lower and upper bonds of the bin. On the y-axis, the event rates for the observations in the bin are shown.

Table 3

Patient scenarios and their resulting estimated probabilities for long-term opioid use.

Scenario	Age	Sex	Language region	Insurance model	Initial opioid dose	Prescriber(s)	Prior episode	Comorbidities	Comedication	Probability for long-term opioid use
1	65	M	Italian	Managed care (Semi) private insurance	30 mg	1	No	None	No	0.44%
2	70	F	German	No managed care model No additional insurance	30 mg	1	Within past 6 months	Lung disease Musculoskeletal disease Neurological disease	Bisphosphonate Benzodiazepine	61.9%
3	49	F	German	No managed care model No additional insurance	<20 mg	>1	Within past 6 months	Chronic infection Cardiovascular disease Psychiatric disease Musculoskeletal disease	Benzodiazepine	89.2%
4	85	M	French	Managed care No additional insurance	50 mg	1	Between 6 months and 2 years	Cardiovascular disease Psychiatric disease Musculoskeletal disease	Bisphosphonate	44.7%
5	75	F	German	Managed care No additional insurance	≥100 mg	>1	Within past 6 months	Inflammatory disease Cardiovascular disease Psychiatric disease Musculoskeletal disease Neurological disease	Bisphosphonate Benzodiazepine Nonopioid analgesic	98.2%
6	75	M	Italian	Managed care No additional insurance	≥100 mg	1	No	Cardiovascular disease Musculoskeletal disease Neurological disease	Nonopioid analgesic	30.7%

logistic regression models and random forest models using all risk factors simultaneously, and no variable selection was necessary. A systematic evaluation in the field of clinical prediction models in oncology by Dhiman et al.[7] reported poor methodological conduct of studies using machine learning. By contrast, we aimed at a rigorous internal–external validation approach for the model derivation and validation, for both strategies including logistic regression and the random forest approaches. Subsequently, we aimed for a neutral comparison of model performance across the 2 strategies.

4.2. Implications for clinical practice

The prediction model has several implications for clinical practice. First, at the time of an initial prescription, the beginning of an opioid episode, it is possible to predict with satisfying accuracy whether the episode will become long-term. An external validation of the prediction model should be performed to evaluate the model's performance in new and prospectively collected patients. This may be performed using other insurance data bases or a prospective collection of patients in a clinical setting. Furthermore, the development of new prediction models for long-term opioid use in other insurance companies and healthcare systems, similar to the implementation of the STORM risk scores derived from the Veteran Health Administration population in the United States, could be enhanced.[31] For patients at high risk for long-term opioid use, early goal directed therapies with multimodal approaches should be discussed, and the awareness of the treating physician should be raised. A collaboration between health insurance company, patients, and physicians may thus result in improved outcome and prevent the need for long-term opioid use. An impact study would reveal the net benefit of the proposed prediction model in clinical practice.

5. Conclusions

Our study demonstrated that long-term opioid use can be predicted at the initiation of an opioid prescription episode, with satisfactory accuracy using data routinely collected at a large health insurance company. Traditional statistical methods resulted in higher discriminative ability and similarly good calibration as compared with machine learning approaches.

Conflict of interest statement

The authors have no conflict of interest to declare.

Acknowledgements

Analysis preregistration: The outline of the research project was described in May 2021. The abstract can be found under https://www.biostat.uzh.ch/fileadmin/biostat/user/Master_thesis_abstract___Pain_Medication.pdf. No formal preregistration of the research project was performed.

Author contributions: AS, UH, and TF had full access to all the data. UH and TF take the responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: UH and MW; Acquisition, analysis, and interpretation of data: AS, UH, TF, MD, JB, and MW; Drafting of the manuscript: UH. Critical revision of the manuscript for intellectual content: TF, MD, AS, JB, and MW. Statistical analysis: TF. Supervision: UH and MW. Analytical code transparency and research materials transparency. Data underlying the study belong to a third party and cannot be made publicly available. This study is based on administrative, deidentified insurance claims data handled in compliance with privacy law and regulations. The analyses were performed within the premises of the Department of Health Sciences at the Helsana, and only aggregated data were shared. Aggregated data can be made available on reasonable request after assessment and permission of the data owner (Helsana Insurances Group). Data requests can be made to Department of Health Sciences, Helsana, Zürichstrasse 130, CH-8600 Dübendorf, Switzerland (website: https://www.helsana.ch/de/private, phone: +41 58 340 18 80, email: Gesundheitskompetenz@helsana.ch). The authors confirm that they did not have any special access to these data which other researchers would not have. The statistical analysis code will be made available on request to the corresponding author.

Ethics declarations.

This study was based on administrative deidentified insurance claims data handled in compliance with privacy law and regulations. According to the local ethical committee (Ethical Committee of the Canton Zurich, Switzerland), no institutional review board (IRB) approval was required. The study was conducted following the principles of good clinical practice and in accordance with the Declaration of Helsinki.

Supplemental digital content

Supplemental digital content associated with this article can be found online at http://links.lww.com/PAIN/B902.

Supplemental video content

A video abstract associated with this article can be found on the journal Web site.

Article history:
Received 30 November 2022
Received in revised form 2 June 2023
Accepted 5 June 2023
Available online 28 September 2023

References

[1] Bedene A, Lijfering WM, Arbous MS, Rosendaal FR, Dahan A, van Dorp ELA. Association between prescription opioid use and unplanned intensive care unit admission and mortality in the adult population of The Netherlands: a registry study. Br J Anaesth 2022;129:254–62.
[2] Beliveau CM, McMahan VM, Arenander J, Angst MS, Kushel M, Torres A, Santos GM, Coffin PO. Stimulant use for self-management of pain among safety-net patients with chronic non-cancer pain. Subst Abus 2022;43:179–86.
[3] Burgstaller JM, Held U, Signorell A, Blozik E, Steurer J, Wertli MM. Increased risk of adverse events in non-cancer patients with chronic and high-dose opioid use-A health insurance claims analysis. PLoS One 2020;15:e0238285.
[4] Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med 2000;19:1141–64.
[5] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22.
[6] 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.
[7] Dhiman P, Ma J, Navarro CLA, Speich B, Bullock G, Damen JAA, Hooft L, Kirtley S, Riley RD, Van Calster B, Moons KGM, Collins GS. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. Bmc Med Res Methodol 2022;22:10.
[8] Els C, Jackson TD, Kunyk D, Lappi VG, Sonnenberg B, Hagtvedt R, Sharma S, Kolahdooz F, Straube S. Adverse events associated with medium- and long-term use of opioids for chronic non-cancer pain: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017;10:CD012509.
[9] Eriksen J, Sjogren P, Bruera E, Ekholm O, Rasmussen NK. Critical issues on opioids in chronic non-cancer pain: an epidemiological study. PAIN 2006;125:172–9.
[10] Federal Statistical Office S. Income distribution. In: CaLC Section Income editor, 2020. Available at. https://www.bfs.admin.ch/bfs/en/home/statistics/economic-social-situation-population/economic-and-social-situation-of-the-population/inqualities-income-ditribution/income-distribution.html.
[11] Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, Steyerberg EW, Collaborators C-T. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol 2020;122:95–107.
[12] Haller E, Watzke B, Blozik E, Rosemann T, Reich O, Huber CA, Wolf M. Antidepressant prescription practice and related factors in Switzerland: a cross-sectional analysis of health claims data. BMC Psychiatry 2019;19:196.
[13] Hu B, Palta M, Shao J. Properties of R-2 statistics for logistic regression. Stat Med 2006;25:1383–95.
[14] Huber CA, Nagler M, Rosemann T, Blozik E, Napflin M, Markun S. Trends in micronutrient laboratory testing in Switzerland: a 7-year retrospective analysis of healthcare claims data. Int J Gen Med 2020;13:1341–8.
[15] Huber CA, Schneeweiss S, Signorell A, Reich O. Improved prediction of medical expenditures and health care utilization using an updated chronic disease score and claims data. J Clin Epidemiol 2013;66:1118–27.
[16] Humphreys K, Shover CL, Andrews CM, Bohnert ASB, Brandeau ML, Caulkins JP, Chen JH, Cuellar MF, Hurd YL, Juurlink DN, Koh HK, Krebs EE, Lembke A, Mackey SC, Larrimore Ouellette L, Suffoletto B, Timko C. Responding to the opioid crisis in north America and beyond: recommendations of the stanford-lancet commission. Lancet 2022;399:555–604.
[17] Kaye AD, Jones MR, Kaye AM, Ripoll JG, Jones DE, Galan V, Beakley BD, Calixto F, Bolden JL, Urman RD, Manchikanti L. Prescription opioid abuse in chronic pain: an updated review of opioid abuse predictors and strategies to curb opioid abuse (Part 2). Pain Physician 2017;20:S111–S133.
[18] Klimas J, Gorfinkel L, Fairbairn N, Amato L, Ahamad K, Nolan S, Simel DL, Wood E. Strategies to identify patient risks of prescription opioid addiction when initiating opioids for pain A systematic review. Jama Netw Open 2019;2:10.
[19] Kuo RN, Dong YH, Liu JP, Chang CH, Shau WY, Lai MS. Predicting healthcare utilization using a pharmacy-based metric with the WHO's anatomic therapeutic chemical algorithm. Med Care 2011;49:1031–9.
[20] Madras BK. The surge of opioid use, addiction, and overdoses responsibility and response of the US health care system. Jama Psychiat 2017;74:441–2.
[21] Mallon T, Ernst A, Brettschneider C, Konig HH, Luck T, Rohr S, Weyerer S, Werle J, Mosch E, Weeg D, Fuchs A, Pentzek M, Kleineidam L, Heser K, Riedel-Heller S, Maier W, Wiese B, Scherer M, Grp AS, Grp AS. Prevalence of pain and its associated factors among the oldest-olds in different care settings - results of the AgeQualiDe study. Bmc Fam Pract 2018;19:10.
[22] Martin BC, Fan MY, Edlund MJ, DeVries A, Braden JB, Sullivan MD. Long-term chronic opioid therapy discontinuation rates from the TROUP study. J Gen Intern Med 2011;26:1450–7.

[23] Miller GC, Valenti L, Britt H, Bayram C. Drugs causing adverse events in patients aged 45 or older: a randomised survey of Australian general practice patients. BMJ Open 2013;3:e003701.
[24] Oh G, Abner EL, Fardo DW, Freeman PR, Moga DC. Patterns and predictors of chronic opioid use in older adults: a retrospective cohort study. Plos One 2019;14:10.
[25] Oliva EM, Bowe T, Tavakoli S, Martins S, Lewis ET, Paik M, Wiechers I, Henderson P, Harvey M, Avoundjian T, Medhanie A, Trafton JA. Development and applications of the Veterans health administration's stratification tool for opioid risk mitigation (STORM) to improve opioid safety and prevent overdose and suicide. Psychol Serv 2017;14:34–49.
[26] Panczak R, Berlin C, Voorpostel M, Zwahlen M, Egger M. The Swiss neighbourhood index of socioeconomic position: update and re-validation. Swiss Med Wkly 2023;153:40028.
[27] Pensa MA, Galusha DH, Cantley LF. Patterns of opioid prescribing and predictors of chronic opioid use in an industrial cohort, 2003 to 2013. J Occup Environ Med 2018;60:457–61.
[28] Putnam KG, Buist DSM, Fishman P, Andrade SE, Boles M, Chase GA, Goodman MJ, Gurwitz JH, Platt R, Raebel MA, Chan KA. Chronic disease score as a predictor of hospitalization. Epidemiology 2002;13:340–6.
[29] R Development Core Team R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2022.
[30] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, Moons KGM, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
[31] Rogal SS, Chinman M, Gellad WF, Mor MK, Zhang HW, McCarthy SA, Mauro GT, Hale JA, Lewis ET, Oliva EM, Trafton JA, Yakovchenko V, Gordon AJ, Hausmann LRM. Tracking implementation strategies in the randomized rollout of a Veterans Affairs national opioid risk management initiative. Implement Sci 2020;15:10.
[32] Schmidt-Hansen M, Bennett MI, Hilgart J. Oxycodone for cancer pain in adult patients. J Am Med Assoc 2015;314:1282–3.
[33] Seyler T, Giraudon I, Noor A, Mounteney J, Griffiths P. Is Europe facing an opioid epidemic: what does European monitoring data tell us? Eur J Pain 2021;25:1072–80.
[34] Sullivan MD, Howe CQ. Opioid therapy for chronic pain in the United States: promises and perils. PAIN 2013;154(suppl 1):S94–S100.
[35] Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167–6.
[36] Von Korff M, Kolodny A, Deyo RA, Chou R. Long-term opioid therapy reconsidered. Ann Intern Med 2011;155:325–8.
[37] Von Korff M, Saunders K, Thomas Ray G, Boudreau D, Campbell C, Merrill J, Sullivan MD, Rutter CM, Silverberg MJ, Banta-Green C, Weisner C. De facto long-term opioid therapy for noncancer pain. Clin J Pain 2008;24:521–7.
[38] Wertli MM, Reich O, Signorell A, Burgstaller JM, Steurer J, Held U. Changes over time in prescription practices of pain medications in Switzerland between 2006 and 2013: an analysis of insurance claims. BMC Health Serv Res 2017;17:167.
[39] Wiener ALaM. Classification and regression by randomForest. R News 2002;2:18–22.
[40] Yarborough BJH, Stumbo SP, Stoneburner A, Smith N, Dobscha SK, Deyo RA, Morasco BJ. Correlates of benzodiazepine use and adverse outcomes among patients with chronic pain prescribed long-term opioid therapy. Pain Med 2019;20:1148–55.
[41] Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. PAIN 2022;163:e328–e332.