Epilepsy & Behavior 89 (2018) 118–125

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Predicting drug-resistant epilepsy — A machine learning approach based on administrative claims data

Sungtae An [a], Kunal Malhotra [a], Cynthia Dilley [b], Edward Han-Burgess [b], Jeffrey N. Valdez [a], Joseph Robertson [b], Chris Clark [b], M. Brandon Westover [c,1], Jimeng Sun [a,*,1]

[a] Georgia Institute of Technology, College of Computing, Atlanta, GA, USA
[b] UCB Pharma, Smyrna, GA, USA
[c] Massachusetts General Hospital, Department of Neurology, Boston, MA, USA

ARTICLE INFO

Article history:
Received 24 July 2018
Revised 4 October 2018
Accepted 8 October 2018
Available online xxxx

Keywords:
Pharmacoresistant
Algorithm
Predictive
Big data

ABSTRACT

Patients with drug-resistant epilepsy (DRE) are at high risk of morbidity and mortality, yet their referral to specialist care is frequently delayed. The ability to identify patients at high risk of DRE at the time of treatment initiation, and to subsequently steer their treatment pathway toward more personalized interventions, has high clinical utility. Here, we aim to demonstrate the feasibility of developing algorithms for predicting DRE using machine learning methods. Longitudinal, intersected data sourced from US pharmacy, medical, and adjudicated hospital claims from 1,376,756 patients from 2006 to 2015 were analyzed; 292,892 met inclusion criteria for epilepsy, and 38,382 were classified as having DRE using a proxy measure for drug resistance. Patients were characterized using 1270 features reflecting demographics, comorbidities, medications, procedures, epilepsy status, and payer status. Data from 175,735 randomly selected patients were used to train three algorithms and from the remainder to assess the trained models' predictive power. A model with only age and sex was used as a benchmark. The best model, random forest, achieved an area under the receiver operating characteristic curve (95% confidence interval [CI]) of 0.764 (0.759, 0.770), compared with 0.657 (0.651, 0.663) for the benchmark model. Moreover, predicted probabilities for DRE were well-calibrated with the observed frequencies in the data. The model predicted drug resistance approximately 2 years before patients in the test dataset had failed two antiepileptic drugs (AEDs). Machine learning models constructed using claims data predicted which patients are likely to fail ≥3 AEDs and are at risk of developing DRE at the time of the first AED prescription. The use of such models can ensure that patients with predicted DRE receive specialist care with potentially more aggressive therapeutic interventions from diagnosis, to help reduce the serious sequelae of DRE.

1. Introduction

Epilepsy is a highly complex disease, or more precisely, a spectrum of diseases associated with different seizure types, a range of severities, and many causes [1–3]. Furthermore, diagnosis is often made entirely on the basis of patient history, leading frequently to misdiagnosis [4,5]. For people with diagnosed epilepsy, antiepileptic drugs (AEDs) are the mainstay of treatment, and most will respond to these drugs and achieve seizure control [1]. However, up to 40% will continue to experience seizures despite treatment, and are classified as having drug-resistant epilepsy (DRE) [6–9]. According to the International League Against Epilepsy, DRE is defined as the failure to achieve sustained seizure freedom after adequate trials of two tolerated and appropriately chosen and used AED schedules (whether as monotherapies or in combination) [9].

Patients with DRE constitute the major burden of epilepsy in the population. Poor seizure control and an escalating drug burden have enormous adverse physical, psychological, and social consequences for patients [7,8,10]. Drug-resistant epilepsy is also associated with an increased risk of mortality; results of a recent, large-scale Canadian study (N = 10,661) showed that 12% of patients died within 2 years of receiving a diagnosis of DRE [11]. Therefore, identifying predictive factors of DRE has been an important area of research for many years. Predicting DRE at diagnosis can help steer the treatment pathway of high-risk patients toward more individualized and specialized interventions, potentially reducing the significant morbidity and mortality associated with DRE. However, evidence for the various identified

* Corresponding author at: Georgia Institute of Technology, 266 Ferst Drive, Atlanta, GA 30313, USA.
E-mail address: jsun@cc.gatech.edu (J. Sun).
[1] Cosenior authors.

https://doi.org/10.1016/j.yebeh.2018.10.013
1525-5050/

factors has not been consistent [10,12]. The use of large volumes of data and advanced, powerful computational techniques could potentially help overcome limitations of standard clinical prediction models, where relatively few (e.g., <100) variables are preselected as potential predictors and even fewer (e.g., <10) are retained in a final prediction model [13].

In the study reported here, the feasibility of using three different machine learning models – Multivariate Logistic Regression, Linear Support Vector Machine, and Random Forest – to predict the probability of DRE at diagnosis was evaluated using claims data from a large cohort of patients with epilepsy with a follow-up time of 10 years.

2. Methods

2.1. Data source and preparation

Comprehensive US claims data spanning a period of 10 years (January 2006–December 2015) were obtained from IQVIA (formerly QuintilesIMS). Claims data are collected routinely for administrative purposes — they are integrated fully from pharmacy, hospitals, and medical clinics at anonymized patient levels. Data are available from very large numbers of patients and allow longitudinal evaluation of healthcare resource use and identification of patterns based on information about diagnoses, procedures, and prescriptions [14]. Data are payer and provider agnostic, indicating that they are not derived from a specific payer or provider but aggregated from multiple systems.

2.2. Patient cohort

Characteristics used to identify patients with epilepsy in the database are summarized in Table 1. Overall, 23 AEDs, older and newer, were included in the analysis, with three classified as rescue medications: diazepam, clonazepam, and lorazepam (Table s1). While clonazepam is also used for chronic therapy, 69% of cases in the dataset appeared to be for short-term use, indicating rescue therapy.

For inclusion in the dataset, patients had to be ≥16 years of age at the time of the first AED prescription and fulfill the criteria of epilepsy diagnosis, data completeness, and drug-resistance. For epilepsy diagnosis, three conditions had to be met: first, having ≥1 epilepsy diagnosis claim with International Classification of Diseases, Ninth Revision (ICD-9) code 345.* or International Classification of Diseases, Tenth Revision (ICD-10) code G40.*, or 2 claims for convulsions, including ICD-9 code 780.39 or ICD-10 code R56.9; second, having ≥1 claim for an AED; and third, the first AED was prescribed as monotherapy with at least 60 days of supply. For the first and second criteria, either a formal epilepsy code OR a convulsion code plus an AED prescription was used as the proxy to identify patients with epilepsy [15]. The third criterion helped ensure that the AED was for long-term treatment rather than an interim prescription until the patient saw another physician (e.g., epileptologist). In terms of data completeness, only patients with ≥1 year of data before receiving their first AED prescription and with active quarterly pharmacy claims every year of their full medical history were included. This prevented inclusion of patients who appeared to be at high risk of developing DRE but were nonadherent with therapy or had missing pharmacy claims data. Finally, only patients classified as being either at high or not-high risk of developing DRE according to study definitions were included.

Table 1
Cohort characteristics.

Characteristic	Value
Patients with epilepsy, N	1,376,756
Median age at index date, years (25th, 75th percentiles)	46.4 (27.0, 61.3)
Female sex, %	54.5
Pharmacy claims, N	28,403,939
Diagnosis claims, N	173,570,273
Inpatient encounters, N	201,202
Outpatient encounters, N	707,332
Emergency room encounters, N	432,915

2.3. Study definitions

Data required for modeling are not always available in claims databases in the format required; therefore, informed assumptions are required in certain cases. The rationale for these assumptions and the definitions used for modeling are provided below (and Supplementary Material).

2.3.1. Drug-resistant epilepsy

Since seizure status (seizure freedom vs. uncontrolled seizures) cannot be captured directly in claims data, proxies for treatment success/failure were used. Success was defined as treatment stability for ≥1 year — if the patient consistently filled the same AED prescription(s), did not require rescue medication for breakthrough seizures, or addition of another AED, the implication is that the patient had at least adequate, if not optimal seizure control. While there may be many reasons for changing, or discontinuing an AED other than lack of efficacy, evidence from a large-scale observational study indicated that after terminal remission, lack of efficacy was indeed the most frequent reason [16]. Conversely, treatment failure was considered to have occurred if a prescription for a different AED was filled, either as a replacement or addition to the current AED, or as a rescue medication.

A proxy was also used for defining DRE — the number of distinct AED schedules prescribed over time [17,18]. The International League Against Epilepsy has defined DRE as failure of two trials of tolerated and appropriately used AED schedules to achieve sustained seizure freedom [9]. However, in our analysis, we considered patients to be at high risk of DRE if they failed three AEDs during the time covered in the dataset. Given that our study was based on predictive, not descriptive, analyses, the use of three AED failures was deemed mathematically appropriate to ensure optimal separation between patients with and without high risk of DRE. Patients with treatment success were considered to have non-DRE, while those who failed one or two AEDs during the time covered in the dataset were considered to be in an indeterminate state and excluded from the analysis.

2.3.2. Time of prediction (index date) and observation and evaluation windows

The 'index date', defined as the time of the prescription of the first AED, was also the time at which DRE is to be predicted (Fig. 1). It was identified by scanning longitudinal records until reaching the first record of an AED. The index date also served as a dividing point in a patient's medical history (Fig. 1); the period before the index date (minimum requirement of ≥1 year before) was the observation window, and the period after, the evaluation window, extended to the patient's last available record (minimum requirement of ≥1 year after for control patients). The prediction made at the index date was based exclusively on data available from the observation window, leading up to and including the index date. Data from the evaluation window were not used in making predictions, but used exclusively to judge the accuracy of predictions. Patients here constitute a mixed population and cannot be categorized as purely case (patients with DRE) and control (patients with non-DRE); given the limited observation period, it is entirely possible that some 'control' patients risk developing DRE after this period. For example, some patients that were false positives for DRE in the model could indeed go on to develop DRE, while chronically treated patients could erroneously be identified as AED-naïve because of insufficient clinical history or transient cessation of treatment ('drug holiday') during the preindex period. However, this 'noisiness' in the data is acceptable given longitudinal limitations of available data sources.

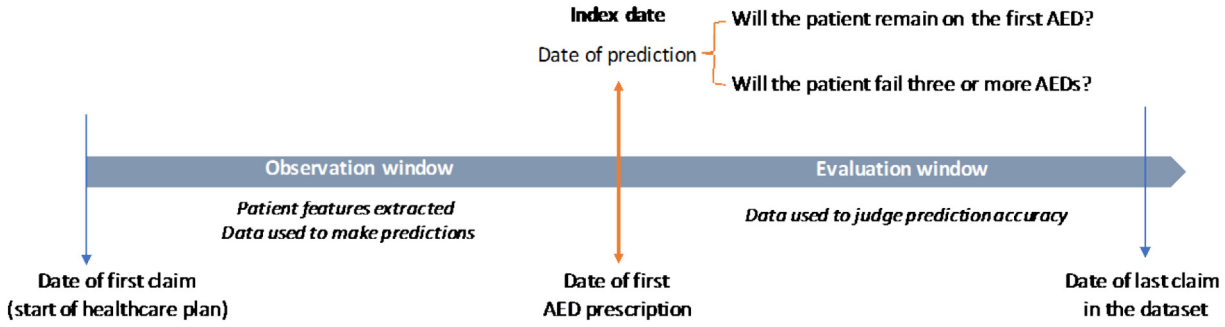

Fig. 1. Experimental setup of the predictive model. Index date, when a prediction is made, is defined as the date of the prescription of the first antiepileptic drug (AED) for each patient. Features used in the models are extracted from the observation window, defined as the period between the index date and the first claim in the dataset. The label for each patient, i.e., whether they are in the case or control group or excluded from the analysis, is determined by identifying how many AED prescriptions are claimed in the evaluation window, defined as the period between the index date and the final claim in the dataset.

2.4. Model building and feature engineering

Three machine learning models typically used for classification were used in this study: Multivariate Logistic Regression (MLR) [19], Linear Support Vector Machine (SVM) [20], and Random Forest (RF) [21]. MLR classifier models have a formulation very similar to simple linear regression, with a linear relationship between the input predictor variables and the logit transformed probability output. Therefore, with this approach, the contribution of individual variables is easy to interpret, making MLR the preferred and best known classifier for clinical prediction models. SVM classifier models leverage higher dimensional space to maximize the distance between two classes in a binary classification, essentially enabling a linearization of the problem in high-dimensional space; while less sensitive to outliers than MLR, SVM may be more computationally intensive given the high-dimensional input space. RF classifier models use an ensemble approach to combine multiple decision trees projected in a randomly chosen subspace (to control overfitting problems associated with decision trees), avoiding reliance on coefficients used in MLR and SVM. RF models are robust to outliers and work well with a high-dimensional input space, nonlinear features, and a space not linearly separable. Given the strengths and weaknesses associated with each model, all three were used to determine the best performing approach.

In the first step, 1270 features, or variables, were extracted from patients' observation window — these fell into five categories: demographics, comorbidities, insurance policy, treatments, and encounters (details in Supplementary Material). Information about the first AED was not included since the model is intended for use at the time when the first AED is being selected.

The dataset was split randomly into three in a 60/20/20 ratio. The largest part constituted the training set, which as the term indicates, was used for training the models to identify patients at risk of DRE. The distribution of patients with DRE in the three datasets — training, validation, and testing — were equivalent, making up 13.13%, 13.05%, and 13.09% of the datasets, respectively. The validation dataset was used for evaluation of model fit on the training dataset during hyperparameter tuning. Hyperparameters define how the model is structured, and the values are set and tuned, or optimized, before training starts. The best hyperparameters were subsequently used to fit the final model. Final model probability outputs for classification from training were calibrated using isotonic regression and fitted to the validation dataset [22]. The testing dataset was used for final model evaluation after hyperparameter tuning, model fit, and probability calibration.

2.5. Assessing prediction accuracy

The trained models were tested by feeding data from the observation window of test patients (testing dataset) to the models for predicting DRE during the evaluation window (Fig. 1). Since a binary classification system was used (patient is or is not at high risk of DRE), Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC; referred to as AUC) and the Average Precision (AP), equivalent to the area under the Precision-Recall (PR) curve, were measured for evaluating model performance. The 95% confidence intervals (CIs) were determined by 10,000 rounds of bootstrapping.

For comparison purposes, a benchmark model was also developed based only on two features: age at onset of epilepsy and sex. To gain qualitative insights into which features contribute most to predictive accuracy, the information gain [23] of features in the RF model was calculated, and a qualitative analysis of the top 20 features with largest information gain was performed. Model calibration, a measure of how well model predictions match the observed DRE frequencies in the data, was assessed by calculating Brier scores and calibration plots [24,25].

3. Results

3.1. Patient cohort

Of 1,376,756 patients in the dataset, 292,892 met the inclusion criteria. Among these, 38,382 (13.10%) had failed three AEDs during

Table 2
Inclusion criteria and their impact on cohort size.

Inclusion criterion	Patients retained N (%)
Epilepsy diagnosis	1,376,756 (100)
– Patient had ≥1 epilepsy diagnosis claim with ICD-9 code 345.* or ICD10 code G40.*, or 2 claims for convulsions, including ICD9 code 780.39 or ICD-10 code R56.9	
– Patient had ≥1 claim for an AED	
– First AED was prescribed as monotherapy with ≥60 days of supply	
Age ≥ 16 years	1,189,582 (86.4)
Data completeness	
Patients with active quarterly pharmacy claims every year	877,105 (63.7)
Drug resistance	
Patients classified according to study definitions as having:[a]	
– Non-DRE (only 1 AED)	532,342 (38.7)
– DRE (failed 3 AEDs)	49,916 (3.6)
Sufficient history	
Patients classified according to study definitions as having:	
– Non-DRE (only 1 AED) + 1-year observation + 1-year evaluation windows	254,510 (18.5)
– DRE (failed 3 AEDs) + 1-year observation window	38,382 (2.8)

DRE = drug-resistant epilepsy.
[a] Patients who failed one or two AEDs during the time covered in the dataset were considered to be in an indeterminate state and excluded.

the evaluation window and were categorized as having DRE, while 254,510 (86.90%) had only one AED prescribed throughout the entire evaluation window and were classified as not being at high risk of DRE (Table 2).

3.2. Model performance — binary predictions

By using 635 features (the top 50%), the mean AUC (95% CI) achieved by each model was 0.764 (0.759, 0.770) for RF, 0.748 (0.742, 0.753) for MLR, and 0.745 (0.740, 0.751) for SVM (Fig. 2A). In comparison, the benchmark model, which used only two features (age at epilepsy onset and sex), achieved an AUC of 0.657 [0.651, 0.663]. The greater the value of the AUC, the better the performance of the model, or in this case, its predictive ability; an AUC of 1 indicates a perfectly accurate prediction while an AUC of 0.5 indicates a prediction based purely on chance.

To assess the impact of the number of features used, numbers were varied from 1% to 100% of 1270 candidate features, chosen by an analysis of variance (ANOVA) F-value test [26] that provides a univariate measure of dependence between target values and each feature (Figure s2). Correspondingly, the mean AP achieved by each model was 0.366 (0.355, 0.377) for RF, 0.335 (0.324, 0.345) for MLR, and 0.327 (0.316, 0.337) for SVM (Fig. 2B). The low AP overall and PR curve could be explained by a highly skewed data distribution (i.e., the number of negative, or non-DRE events outnumbering the positive, or DRE, events). Precision compares false positives to true positives rather than true negatives (as is the case with ROC analysis), so a large number of negative (non-DRE) events would have a significant impact on the PR curve while having little impact on AUC [27]. Nonetheless, the predictive models, especially RF, substantially outperformed the benchmark model in terms of both AUC and AP.

To further assess the calibration of the prediction models, Brier scores were calculated. The Brier score provides a measure of the accuracy of the prediction [24]. It ranges from 0 to 1, with lower scores indicating better calibrated predictions — the best possible score is 0, for a totally accurate prediction [28]. In calibration reliability plots, the diagonal indicates the performance of a perfectly calibrated prediction; in the case of the current study, the predicted probability of drug resistance matching precisely the observed percentage (Fig. 3). A well-calibrated model would have the curve as close as possible to this diagonal. The Brier scores (95% CI) were 0.099 (0.097, 0.101) for RF, 0.102 (0.100, 0.103) for MLR, and 0.106 (0.104, 0.108) for SVM, indicating better calibration than the benchmark model (0.110; 0.108, 0.112).

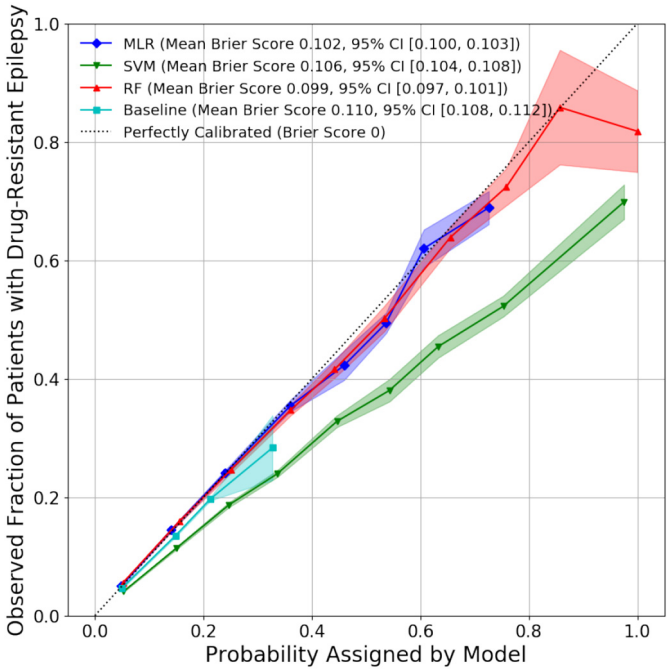

Fig. 3. Risk Calibration curves for each model, generated after the probability calibration step with isotonic regression. The Brier score, which ranges from 0 to 1, is also reported. A curve closer to the dotted diagonal line, with a corresponding lower Brier score, indicates better calibration (MLR = multivariate logistic regression; RF = random forest; SVM = support vector machine).

3.3. Qualitative results

The predictive power of each feature selected during the model training procedure for inclusion in the final model was evaluated. The top 20 features with the highest information gain values fell into three relatively natural groups (Table 3): indicators of a patient's level of

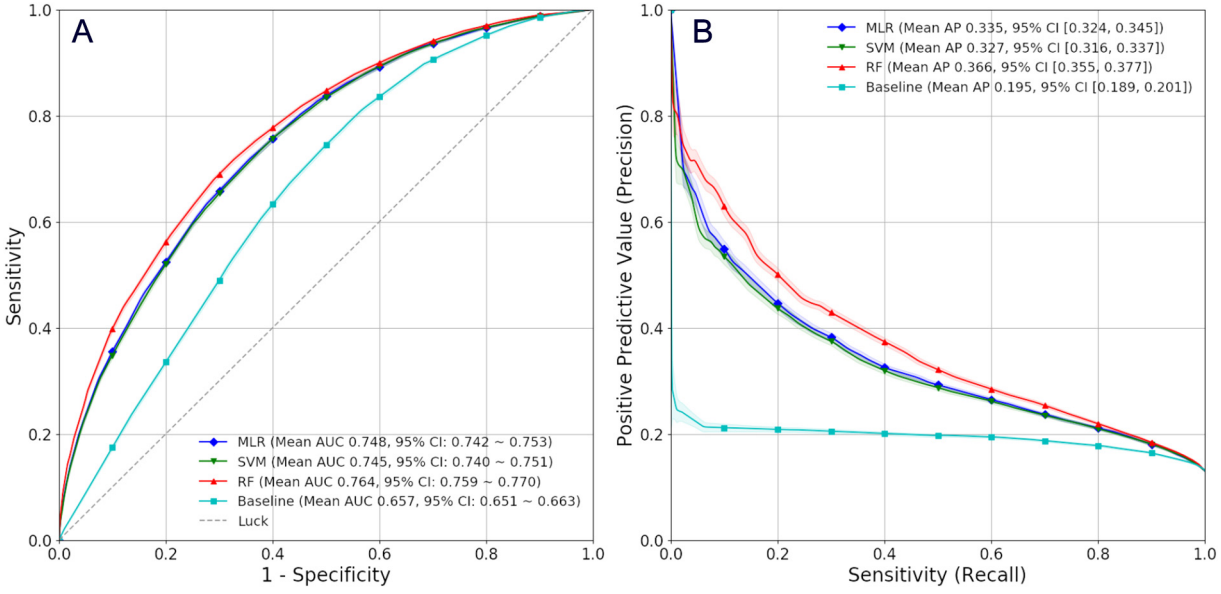

Fig. 2. Evaluation of model accuracy using Receiver Operating Characteristic (A) and Precision-Recall (B) curves. Both types of curves are generated from 10,000 times of bootstrapped results. Solid lines are mean curves and shaded regions are their 95% confidence intervals (MLR = multivariate logistic regression; RF = random forest; SVM = support vector machine).

Table 3
Top 20 predictive features in the best performing model.

Feature
Indicators of comorbidities
Age
Charlson Comorbidity Index score
Number of claims at any past time for lipid-lowering drugs classified as USP Class "Dyslipidemics HMG CoA Reductase Inhibitors"
Presence of claims for diagnosis of hypertensive disease classified as CCS 98 "Essential hypertension" or CCS 99 "Hypertension with complications and secondary hypertension"
Number of claims prior to the past year for diagnosis of essential hypertension classified as CCS 98 "Unspecified essential hypertension"
Indicators of epilepsy and epilepsy complexity
Number of claims in the past year with an epilepsy diagnostic code classified as CCS 83: "Epilepsy; convulsions"
Number of claims prior to the past year with an epilepsy diagnostic code classified as CCS 83: "Epilepsy; convulsions"
Epilepsy comorbidity score
First antiepileptic drug is prescribed by a neurologist (yes/no [1/0])
Indicators of level of activity within the medical system
Number of claims prior to the past year for diagnostic tests classified as CCS 227: "Other diagnostic procedures: interview; evaluation; consultation"
Payer for first AED: Medicaid
Payer for any drug in the past year: Medicaid
Number of claims in the past year for any procedure
Number of months in the past year with claims for any diagnosis
Number of claims in the past year for procedures classified as CCS 227: "Other diagnostic procedures: interview; evaluation; consultation"
Number of months in the past year with claims for any prescription
Number of claims prior to the past year for chemistry and hematology tests classified as CCS 233: "Chemistry and hematology"
Number of claims prior to the past year for therapeutic procedures classified as CCS 231: "Other therapeutic procedures"
Number of claims prior to the past year for disorders of lipid metabolism classified as CCS 53 "Disorders of lipid metabolism"
Payer for first AED: Medicare

CCS, Clinical Classification Software (available at https://www.hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf);
USP drug class, The US Pharmacopeial Convention drug classification (available at https://www.usp.org/sites/default/files/usp/document/our-work/healthcare-quality-safety/uspmmg_v6_0_cat-class.pdf).

activity within the medical system (e.g., frequency of tests, procedures, medical visits); indicators of comorbidities (e.g., age, need for drugs associated with cerebrovascular disease risk); and indicators of disease severity (e.g., frequency of visits associated with an epilepsy diagnosis code).

Potential time saved by the best performing model, RF, relative to the status quo of being classified as having DRE when a patient fails the second AED – for any reason – was also assessed. Fig. 4 shows the survival analysis (Nelson–Aalen) estimate [29,30] of cumulative distribution fraction of patients who have failed two AEDs in the test dataset. The time axis measures the period between the first AED prescription (the index date) and the third AED prescription (the second failure); control cases are treated as censored at the time they drop out of the dataset, since by definition, the third AED prescription was not observed. The mean and median value of time taken to be determined as having DRE (or having cycled through three AEDs) were 1.97 and 1.58 years, respectively. Consequently, the model predicted the presence, or development of DRE on average approximately 2 years before patients failed two AEDs, the point of which they are classified as having DRE.

4. Discussion

The three machine learning models in this study showed potential in identifying patients presenting with DRE at diagnosis. Based on the AUC, which is used to assess the quantitative performance of predictive models, the greatest accuracy was obtained with the RF model (AUC 0.764, 95% CI 0.759, 0.770); predictions were also well-calibrated (Brier score 0.099, 95% CI 0.097, 0.101). Results suggested that this prediction capability can potentially identify patients at high risk of DRE 1.97 years earlier than the current practice of waiting for them to fail two AEDs. It should be noted that longitudinal data limitations meant that patients were tracked for only several years in this study. In the real world however, it may take longer than 2 years for a patient to fail two drugs, in which case early prediction may become even more valuable than indicated here. Limitations of the dataset are also the likely explanation for the lower proportion of the patients with DRE in this study (13.10%) than that observed in the real-world setting; the limited time span covered by the dataset necessitated patients to fail three AEDs within a few years for the failure to be captured.

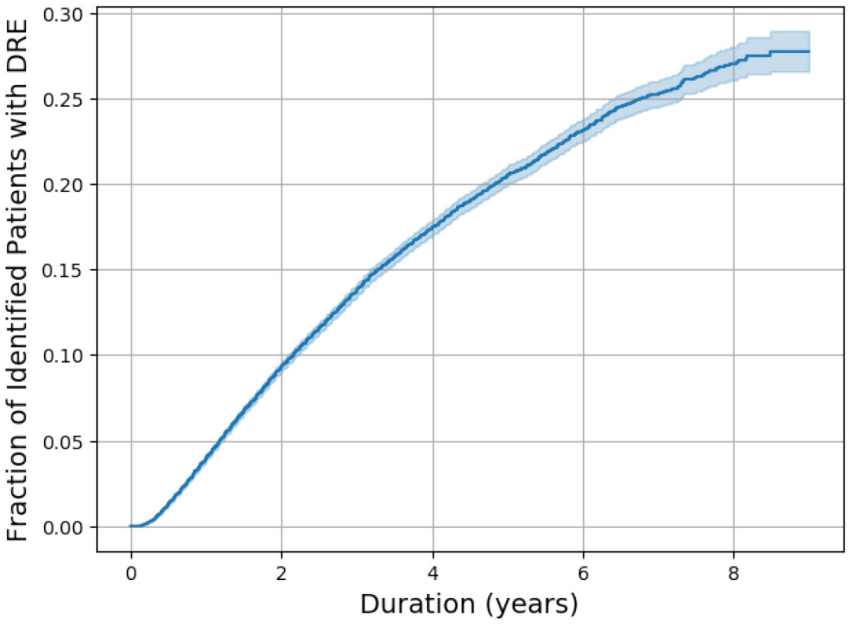

Fig. 4. Nelson–Aalen cumulative hazard function curve for status quo. The x-axis represents the time taken for a given patient to be classified as having drug-resistant epilepsy (DRE) after prescription of the first antiepileptic drug. The corresponding value on the y-axis represents the fraction of patients with DRE classified up to that time, as defined by the status quo.

Many studies have been conducted in the attempt to identify patient characteristics associated with DRE; however, most included relatively small sample sizes. In one of the larger studies, which included 780 patients newly diagnosed as having epilepsy followed-up over 20 years, those reporting more than 10 seizures before treatment initiation were more than twice as likely to develop DRE than those reporting fewer seizures [31]. Logistic regression analyses demonstrated that DRE was also associated with family history of epilepsy, previous febrile seizures, traumatic brain injury, intermittent recreational drug use, and previous or current psychiatric comorbidity, particularly depression [31]. In other smaller-scale studies, conducted in the pediatric population, the most consistent factors predicting DRE were high seizure frequency before treatment and symptomatic etiology [32–34].

For the analyses in the current study, a diverse set of 1270 features from 292,892 patients over an observation period of 10 years was used, leading to the development of models that, within the limits of the methodology of this study, predicted those patients who had, or developed DRE after receiving their first AED. Using a multivariate model to identify key predictors within the dataset, the top 20 features with the highest information gain values were found to be indicators of comorbidities, and indicators of healthcare utilization, which could, in turn be indicative of disease severity. These are similar to the findings of the studies described above, reinforcing the validity of claims data for such proxies.

The seminal study by Kwan and Brodie showed that in some patients, DRE may be present at onset, rather than evolve over time [35]. In others patients though, the condition will evolve and develop at later stages, while some will follow a relapsing–remitting pattern of seizure control [36]. Nonetheless, the ability to identify individuals at high risk of DRE, ideally at the time they are prescribed the first AED, has enormous clinical utility. Indeed, development of a screening test or protocol that could "identify individuals with persistent seizures who need to be referred to an epileptologist for further evaluation and treatment" was one of the research priorities highlighted in the Institute of Medicine's 2012 report, Epilepsy across the spectrum: Promoting health and understanding [3].

The care of patients with epilepsy typically starts with an evaluation in an emergency room or a primary care physician's office [37]. While most patients can be initially evaluated and managed at the first or second level of epilepsy care by a primary care physician or a general neurologist, referral to tertiary care is recommended if seizures persist [37,38]. However, it has been reported that less than 1% of people with DRE are evaluated at a full-service epilepsy center [6]. A recent US National Health Interview Survey revealed that 38% of patients with uncontrolled seizures had not seen a neurologist or epilepsy specialist in the year preceding the survey [39]. In a large-scale survey conducted in the US, France, and Germany, 19% of patients reported persistent seizures, poor quality of life, and low satisfaction with treatment, yet they had not had a treatment change in 2 years [40]. These observations indicate an important treatment gap for patients with DRE who require specialist care. Indeed, in a survey of patients based on the quality measures for epilepsy care developed by the American Academy of Neurology (AAN) [41], referral of patients with DRE to an epileptologist was one of the most important identified gaps [42]. The importance of identifying patients with DRE is highlighted by the results of a large-scale Canadian study (N = 10,661), where 12% of patients with DRE died within 2 years of receiving a DRE diagnosis [11]. In the study, mortality was substantially higher among patients with DRE than among matched control patients without DRE (12% vs. 1.1%; odds ratio 14.6).

The use of a predictive tool can help in closing the treatment gap by identifying patients at high risk of developing DRE — these patients can be monitored more closely and regularly than those who are not, or they can be referred straight away to centers where they can receive care from an interdisciplinary team directed by a neurologist or neurosurgeon with special expertise in epilepsy [37,38]. Here, the management plan of patients can be steered toward a more individualized, aggressive plan; for example, combination AED therapy, palliative interventions such as vagal nerve stimulation, or potentially curative resective surgery may be considered at a much earlier stage [43]. According to the AAN epilepsy care measures, patients with DRE should be assessed for referral to a higher level of care on a regular basis [41]. At an even more basic level, video-electroencephalography in an epilepsy monitoring unit can be performed to rule out misdiagnosis [44], which has been reported to occur in 20–31% of cases [37]. In a survey of physicians, only primary care practitioners and general neurologists reported having misdiagnosed epilepsy, while none of the epileptologists did [40]. A low correlation between referral and specialist diagnosis has also been reported [45].

It is clear that there are numerous potential gains associated with early detection of any risk factor, or disease, notably, prolongation of, and improvement in the quality of life, due to reduction in severity and frequency of a disease state and/or prevention and delay of its complications; it is also associated with significant reductions in healthcare costs [46]. Therefore, in recent years, machine learning techniques have been used increasingly for risk stratification and predicting outcomes in a broad range of clinical settings. For example, using a similar database to that used in the current study (US administrative claims, pharmacy records and laboratory results), Razavian and colleagues identified 42,000 variables from 4.1 million individuals to develop a model that was able to predict the onset of type 2 diabetes at a population level [47]. Similarly, administrative data were also used to develop models for predicting heart failure [48]. Smaller datasets have been used for predictive modeling of the risk of suicide attempt among patients with depression, mortality among patients with cardiovascular disease, or progression of Parkinson's disease and cancer among others [49–52].

The large size of the dataset, and consequently statistical power of the analysis, is among the strengths of this study. Another is the use of an open claims dataset [53], which means patients were linked across multiple payers over time and those who did not stay with the same job/employer for 10 years or had government insurance were still represented. Such open data approximate the average patient population in terms of stability and payer variation, as well as demographics, comorbidities, and payer types across the US. Additionally, open data can be more inclusive of adult patients with DRE who may often switch insurance frequently because of their inability to maintain stable employment. Open datasets can also have limitations; for example, data may be missing if a patient switches to a nonreporting pharmacy. While we attempted to control for continuous eligibility (data reporting), the method is not fully reliable. All studies based on claims data are subject to limitations [54,55]. Seizure frequency is not reported in claims databases, and other relevant information, such as etiology and electroencephalographic data, may also be lacking. This also means that the distinction between patients with controlled epilepsy and those with DRE based on changes in AED use is indirect. Furthermore, identification of patients with epilepsy based on diagnostic codes and AED prescriptions might not always correspond with direct chart review [54]. Nonetheless, while treatment changes may arise from several causes (e.g., lack of efficacy, side effects, nonadherence), it is likely that these correspond reasonably well with AED failures reported in other studies of DRE. Similarly, while treatment stability can be reasonably assumed to signify good seizure control, it is possible that in some cases this simply represents reluctance of physicians or patients to make changes in the treatment schedule. Finally, the features identified are not very specific, making it difficult to develop rules-based screening strategies for the population with DRE specifically. Clinicians rely on a rule-based approach for clinical care; therefore, use of a score-based approach that relies on complex algorithms for DRE classification might be less interpretable and intuitive for clinicians.

5. Conclusion

In the current study, predictive models based on claims data demonstrated potential in identifying patients at high risk of developing DRE at the time of first AED prescription. The best model predicted DRE on average 1.97 years before a patient is assumed to fail two AEDs, the recognized definition of DRE. The use of such models as a clinical decision support system can potentially transform the lives of patients with DRE through early identification and implementation of personalized, more aggressive treatment strategies earlier in the course of their disease. Future work should include a comparison of the performance of the models described here against drug resistance ascertained in prospective clinical studies. A clinical study comparing outcomes among patients following different referral patterns can also be valuable. Finally, future work on more interpretable machine learning models that identify informative features for health outcomes may be necessary before widespread adoption of machine learning techniques in clinical settings.

Conflict of interest

Cynthia Dilley, Edward Han-Burgess, Joseph Robertson, and Chris Clark are employees of UCB Pharma. The remaining authors declare no conflict of interest. The study was funded by UCB Pharma.

Acknowledgments

The authors thank Azita Tofighy for critically reviewing the manuscript, and editorial assistance, which was funded by UCB Pharma. They also thank Barbara Pelgrims, UCB Pharma, for overseeing the development of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yebeh.2018.10.013.

References

[1] Duncan JS, Sander JW, Sisodiya SM, Walker MC. Adult epilepsy. Lancet 2006;367:1087–100.
[2] Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon CS, Dykeman J, et al. Prevalence and incidence of epilepsy: a systematic review and meta-analysis of international studies. Neurology 2017;88:296–303.
[3] Institute of Medicine (IOM). Epilepsy across the spectrum: promoting health and understanding. Washington DC: The National Academies Press; 2012.
[4] Ferrie CD. Preventing misdiagnosis of epilepsy. Arch Dis Child 2006;91:206–9.
[5] Xu Y, Nguyen D, Mohamed A, Carcel C, Li Q, Kutlubaev MA, et al. Frequency of a false positive diagnosis of epilepsy: a systematic review of observational studies. Seizure 2016;41:167–74.
[6] Engle Jr J. What can we do for people with drug-resistant epilepsy? Neurology 2016;87:2483–9.
[7] Laxer KD, Trinka E, Hirsch LJ, Cendes F, Langfitt J, Delanty N, et al. The consequences of refractory epilepsy and its treatment. Epilepsy Behav 2014;37:59–70.
[8] Weaver DF, Pohlmann-Eden B. Pharmacoresistant epilepsy: unmet needs in solving the puzzle(s). Epilepsia 2013;54:80–5.
[9] Kwan P, Arzimanoglou A, Berg AT, Brodie MJ, Allen Hauser W, Mathern G, et al. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. Epilepsia 2010;51:1069–77.
[10] Mohanraj R, Brodie MJ. Early predictors of outcome in newly diagnosed epilepsy. Seizure 2013;22:333–44.
[11] Burneo JG, Shariff SZ, Liu K, Leonard S, Saposnik G, Garg AX. Disparities in surgery among patients with intractable epilepsy in a universal health system. Neurology 2016;86:72–8.
[12] Wassenaar M, Leijten FS, Egberts TC, Moons KG, Uijl SG. Prognostic factors for medically intractable epilepsy: a systematic review. Epilepsy Res 2013;106:301–10.
[13] Groeneveld PW, Rumsfeld JS. Can big data fulfill its promise? Circ Cardiovasc Qual Outcomes 2016;9:679–82.
[14] Stein JD, Lum F, Lee PP, Rich 3rd WL, Coleman AL. Use of health care claims data to study patients with ophthalmologic conditions. Ophthalmology 2014;121:1134–41.
[15] Faught E, Richman J, Martin R, Funkhouser E, Foushee R, Kratt P, et al. Incidence and prevalence of epilepsy among older US Medicare beneficiaries. Neurology 2012;78:448–53.
[16] Giussani G, Bianchi E, Canelli V, Erba G, Franchi C, Nobili A, et al. Antiepileptic drug discontinuation by people with epilepsy in the general population. Epilepsia 2017;58:1524–32.
[17] Devinsky O. Patients with refractory seizures. N Engl J Med 1999;340:1565–70.
[18] Berg AT, Kelly MM. Defining intractability: comparisons among published definitions. Epilepsia 2006;47:431–6.
[19] Bewick V, Cheek L, Ball J. Statistics review 14: logistic regression. Crit Care 2005;9:112–8.
[20] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004;14:199–222.
[21] Denil M, Matheson D, Freitas ND. In: Xing EP, Jebara T, editors. Narrowing the gap: random forests in theory and in practice. Proceedings of the 31st international conference on machine learning. Bejing, China: PMLR; 2014. p. 665–73.
[22] Chakravarti N. Isotonic median regression: a linear programming approach. Math Oper Res 1989;14:303–8.
[23] Michalski RS, Carbonell JG, Mitchell TM. Machine learning: an artificial intelligence approach. Springer Science & Business Media; 2013.
[24] Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950;78:1–3.
[25] Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. Proceedings of the 22nd international conference on machine learning. Bonn, Germany: ACM; 2005. p. 625–32.
[26] Lowry R. Concepts and applications of inferential statistics. Available at http://vassarstats.net/textbook/; 2014, Accessed date: September 2018.
[27] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. Proceedings of the 23rd international conference on machine learning, New York, NY, USA. ACM; 2006. p. 233–40.
[28] Steyerberg EW, Harrell Jr FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–81.
[29] Nelson W. Theory and applications of hazard plotting for censored failure data. Dent Tech 1972;14:945–66.
[30] Aalen O. Nonparametric inference for a family of counting processes. Ann Stat 1978:701–26.
[31] Hitiris N, Mohanraj R, Norrie J, Sills GJ, Brodie MJ. Predictors of pharmacoresistant epilepsy. Epilepsy Res 2007;75:192–6.
[32] Sillanpää M, Schmidt D. Early seizure frequency and aetiology predict long-term medical outcome in childhood-onset epilepsy. Brain 2009;132:989–98.
[33] Gururaj A, Sztriha L, Hertecant J, Eapen V. Clinical predictors of intractable childhood epilepsy. J Psychosom Res 2006;61:343–7.
[34] Kwong KL, Sung WY, Wong SN, So KT. Early predictors of medical intractability in childhood epilepsy. Pediatr Neurol 2003;29:46–52.
[35] Kwan P, Brodie MJ. Early identification of refractory epilepsy. N Engl J Med 2000;342:314–9.
[36] Brodie MJ, Barry SJ, Bamagous GA, Norrie JD, Kwan P. Patterns of treatment response in newly diagnosed epilepsy. Neurology 2012;78:1548–54.
[37] Labiner DM, Bagic AI, Herman ST, Fountain NB, Walczak TS, Gumnit RJ, et al. Essential services, personnel, and facilities in specialized epilepsy centers—revised 2010 guidelines. Epilepsia 2010;51:2322–33.
[38] National Institute for Health and Clinical Excellence. The epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care. Clinical guideline 137; 2012.
[39] Thurman DJ, Kobau R, Luo YH, Helmers SL, Zack MM. Health-care access among adults with epilepsy: the U.S. National Health Interview Survey, 2010 and 2013. Epilepsy Behav 2016;55:184–8.
[40] Groenewegen A, Tofighy A, Ryvlin P, Steinhoff BJ, Dedeken P. Measures for improving treatment outcomes for patients with epilepsy—results from a large multinational patient–physician survey. Epilepsy Behav 2014;34:58–67.
[41] Fountain NB, Van Ness PC, Swain-Eng R, Tonn S, Bever Jr CT, American Academy of Neurology Epilepsy Measure Development Panel and the American Medical Association-Convened Physician Consortium for Performance Improvement Independent Measure Development Process. Quality improvement in neurology: AAN epilepsy quality measures. Neurology 2011;76:94–9.
[42] Wicks P, Fountain NB. Patient assessment of physician performance of epilepsy quality-of-care measures. Neurol Clin Pract 2012;2:335–42.
[43] Kwan P, Schachter SC, Brodie MJ. Drug-resistant epilepsy. New Engl J Med 2011;365:919–26.
[44] Shih JJ, Fountain NB, Herman ST, Bagic A, Lado F, Arnold S, et al. Indications and methodology for video-electroencephalographic studies in the epilepsy monitoring unit. Epilepsia 2018;59:27–36.
[45] Angus-Leppan H. Diagnosing epilepsy in neurology clinics: a prospective study. Seizure 2008;17:431–6.
[46] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104–16.
[47] Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. Big Data 2015;3:277–87.
[48] Ng K, Steinhubl SR, deFilippi C, Dey S, Stewart WF. Early detection of heart failure using electronic health records: practical implications for time before diagnosis, data diversity, data quantity, and data density. Circ Cardiovasc Qual Outcomes 2016;9:649–58.
[49] Passos IC, Mwangi B, Cao B, Hamilton JE, Wu MJ, Zhang XY, et al. Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. J Affect Disord 2016;193:109–16.
[50] Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected

coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J 2017;38:500–7.
[51] Latourelle JC, Beste MT, Hadzi TC, Miller RE, Oppenheim JN, Valko MP, et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. Lancet Neurol 2017;16:908–16.
[52] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2014; 13:8–17.
[53] Kee VR, Gilchrist B, Granner MA, Sarrazin NR, Carnahan RM. A systematic review of validated methods for identifying seizures, convulsions, or epilepsy using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:183–93.
[54] Moura LM, Price M, Cole AJ, Hoch DB, Hsu J. Accuracy of claims-based algorithms for epilepsy research: revealing the unseen performance of claims-based studies. Epilepsia 2017;58:683–91.
[55] Devinsky O, Dilley C, Ozery-Flato M, Aharonov R, Goldschmidt Y, Rosen-Zvi M, et al. Changing the approach to treatment choice in epilepsy using big data. Epilepsy Behav 2016;56:32–7.