Target Adverse Event Profiles for Predictive Safety in the Postmarket Setting

Peter Schotland[1,4], Rebecca Racz[1], David B. Jackson[2], Theodoros G. Soldatos[2], Robert Levin[3], David G. Strauss[1] and Keith Burkhart[1,*]

We improved a previous pharmacological target adverse-event (TAE) profile model to predict adverse events (AEs) on US Food and Drug Administration (FDA) drug labels at the time of approval. The new model uses more drugs and features for learning as well as a new algorithm. Comparator drugs sharing similar target activities to a drug of interest were evaluated by aggregating AEs from the FDA Adverse Event Reporting System (FAERS), FDA drug labels, and medical literature. An ensemble machine learning model was used to evaluate FAERS case count, disproportionality scores, percent of comparator drug labels with a specific AE, and percent of comparator drugs with the reports of the event in the literature. Overall classifier performance was F1 of 0.71, area under the precision-recall curve of 0.78, and area under the receiver operating characteristic curve of 0.87. TAE analysis continues to show promise as a method to predict adverse events at the time of approval.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ A prior pilot study of six drugs demonstrated that pharmacological target adverse event (TAE) profiles based on marketed drugs can be used to predict unlabeled adverse events (AEs) for a new drug at the time of approval.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ Can machine learning techniques applied to target AE profiles predict unlabeled adverse events at the time of new drug approval in a larger set of drugs?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ A machine learning model that used data from the US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS), peer-reviewed literature and FDA drug labels for comparator drugs that bind similar targets was able to predict postmarket AEs.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ This approach may improve postmarket pharmacovigilance by being able to focus resources on predicted AEs. Additionally, this approach can be applied at any stage of drug development.

Many adverse events (AEs; adverse drug reactions) are identified in the postmarketing period and often undergo a costly, time-consuming analysis before a safety label change or other regulatory decision is made related to a product.[1] The US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) MedWatch reporting has increased to over 1.8 million reports per year. Automated tools that provide mechanistic insights and signal strengthening are needed to identify rare AEs for augmented pharmacovigilance. Efforts to predict AEs have utilized a variety of data sources, including FAERS reports,[2–5] drug labels,[2,4,6] signaling pathways,[7] chemical features,[7,8] gene expression,[8] literature,[6] the electronic health record,[9] prescription records,[10] and social media.[11] Several of these algorithms have demonstrated excellent performance. For example, a machine learning algorithm utilizing multiple chemical and biological features achieved a precision of 66% and successfully predicted AEs associated with several drug withdrawals.[7] Additionally, an ensemble method was used to identify adverse drug events on social media datasets, achieving area under the receiver operating characteristics curve values of about 80%.[11] Knowledge gained from these data mining analytics will also provide important safety information for drug development.[12]

Our previous pilot work created a model based on data from FAERS and drug labels.[4] Molecular target adverse event (TAE)

[1]Division of Applied Regulatory Science, Office of Clinical Pharmacology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [2]Molecular Health GmbH, Heidelberg, Germany; [3]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [4]Present address: Office of Oncologic Diseases, Office of New Drugs, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA. *Correspondence: Keith K. Burkhart (keith.burkhart@fda.hhs.gov)

Received September 20, 2019; accepted August 31, 2020. doi:10.1002/cpt.2074

profiles were created by the selection of comparator drugs that closely resemble the target activity of the drug of interest. Our previous work achieved a precision of 0.67, recall of 0.81, and specificity of 0.71. In this report, we have added literature reports as another data source as well as additional features for learning. We tested an ensemble learning method to predict unlabeled AEs using data available at the time of drug approval.

METHODS

Overview

The purpose of this study is to assess the ability of an ensemble machine learning model with features constructed from TAE profiles to predict drug AEs (adverse drug reactions) in the postmarket setting. We created an ensemble of low-complexity classification methods, described in detail below to predict AEs found on the product labels of FDA approved drugs with at least 3 years postmarket exposure using only data available at the time of approval. TAEs are generated by aggregating AE data for a set of comparator drugs sharing pharmacological receptors with a drug of interest. Three data sources are used to generate TAEs: FAERS reports for comparator drugs, literature reports for comparator drugs, and FDA product labels of comparator drugs. Retrospective data are used for TAE/feature generation as the use of postmarket data may add an optimistic bias to model performance. Predictions are restricted to a list of select AEs, termed designated medical events (DMEs), chosen in consultation with medical officers in the Office of Surveillance and Epidemiology at the FDA. See **Figure 1** for an overview of the workflow. See TA_55_drug_study.zip in **Supplementary Materials** for the full list of study drugs and their comparator drugs.

Study drugs

We used all prescription drugs approved by the FDA between January 2008 and December 2013 that are not first-in-class and are not combination products (i.e., have more than one active moiety). January 2008 ensures product labels are in Structured Product Labeling format,[13] allowing for accurate text-mining of safety content (described below). Fifty-seven drugs were identified, two of which (canakinumab and ospemifene) were excluded from this study because of a lack of sufficient FAERS data for their comparator drugs. See **Table S1** for the list of drugs. A more detailed list, including comparator drugs used to generate TAE profiles can be found in the **Supplementary Materials**. One drug, dapagliflozin, approved January 8, 2014, was included in error but retained in the final dataset.

Designated medical events

There are over 20,000 preferred terms (PTs) in the Medical Dictionary for Regulatory Activities (MedDRA)[14] and roughly 4,000 MedDRA PTs were identified in the FDA prescription drug labels by text-mining. To reduce the number of potential AEs in our model to a manageable level, we chose 167 MedDRA PTs grouped into 36 AE categories, termed DMEs. Each DME is defined by a group of MedDRA PTs and no PT was used more than once per DME.

The DME list was constructed in consultation with medical officers at the FDA and represents a broad range of drug AEs of interest to the FDA, including rare AEs often undetected during premarket evaluation. The grouping of related MedDRA PTs is a common practice at the FDA that helps address the granularity of MedDRA. For example, the terms "cerebral haemorrhage" and "cerebrovascular accident" may be used by different reporters to refer to the same AE. Thus, combining PTs into DMEs allows the aggregation of AE information to capture relevant medical events with similar etiology and/or clinical significance.

The DME list was constructed before dataset creation and model building. DMEs on drug labels were identified using text-mining and manual curation. See **Table 1** for the full list of DMEs and their prevalence (proportion of study drugs labeled for a DME) in our data set.

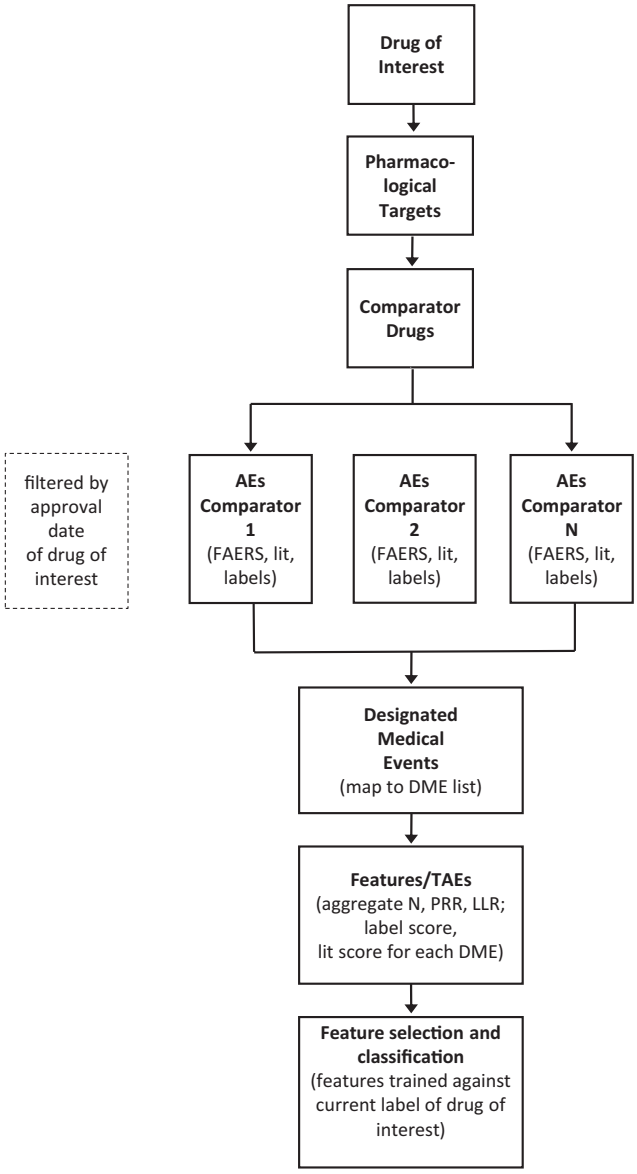

Figure 1 Target analysis workflow. AEs, adverse events; FAERS, USA Food and Drug Administration (FDA) Adverse Event Reporting System; LLR, Log likelihood ratio; N, associated case count; PRR, Proportional Reporting Ratio; TAEs, target adverse-event.

Generation of target adverse-event profiles

Target-adverse event profiles from FAERS reports. TAEs from FAERS reports were generated using a bioinformatics tool, EFFECT from Molecular Health, GmbH.[15] Proportional Reporting Ratios (PRRs) are calculated using the approach described by van Puijenbroek *et al.*[16] and Evans *et al.*[17] and as shown in the example equation below:

$$\mathrm{PRR} = \frac{\text{cases of desvenlafaxine and serotonin syndrome} * \text{all desvenlafaxine cases}}{\text{all other serotonin syndrome cases} * \text{all events for all other drugs}}$$

Log likelihood ratio (LLR) and adjusted P value are calculated according to Huang, *et al.*[18] Analogous to the computation of PRR for drug-AE pairs, the software computes PRR and LLR for target-AE pairs. EFFECT uses DrugBank to identify

Table 1 Designated medical events

Designated medical event	MedDRA PT
Abnormal bleeding	Cerebellar haemorrhage, cerebral haemorrhage, coagulopathy, gastrointestinal haemorrhage, haematoma, haemorrhage, haemorrhage intracranial, rectal haemorrhage, vaginal haemorrhage
Accidents and injuries	Accident, fall, fracture, injury, paralysis, road traffic accident
Acute and chronic pancreatitis	Pancreatitis, pancreatitis acute
Autoimmunity	Haemolytic anaemia, myositis, vasculitis
Bone marrow failure	Anaemia, aplastic anaemia, bone marrow failure, pancytopenia, thrombocytopenia
Arterial thrombotic event	Acute myocardial infarction, angina pectoris, cerebrovascular accident, myocardial infarction, transient ischaemic attack
Cardiac arrhythmia	Arrhythmia, atrial fibrillation, atrioventricular block, bradycardia, supraventricular tachycardia, tachycardia, ventricular arrhythmia, ventricular extrasystoles, ventricular fibrillation, ventricular tachycardia
Colitis excl infective	Colitis, colitis ulcerative, Crohn's disease
Deliria	Aggression, amnesia, confusional state, delirium, delusion, disorientation, hallucination, hostility, memory impairment, paranoia
Edema	Oedema, oedema peripheral
Extrapyramidal symptoms	Abasia, akathisia, dyskinesia, dystonia, extrapyramidal disorder, hyperkinesia, hypertonia, tardive dyskinesia
Heart failure	Cardiac failure, cardiac failure congestive, cardiomyopathy, pulmonary oedema
Hepatic toxicity	Cholestasis, hepatic failure, hepatic necrosis, hepatitis, hepatotoxicity, jaundice, jaundice cholestatic, liver injury
Hypersensitivity	Anaphylactic reaction, anaphylactic shock, anaphylactoid reaction, angioedema, eosinophilia, hypersensitivity, laryngeal oedema, photosensitivity reaction, urticaria
Hypertension	Hypertension
Impaired wound healing	Gastric ulcer, impaired healing, peptic ulcer, skin ulcer, stomatitis, ulcer
Infection and infestation	Bacterial infection, bronchitis, candida infection, cellulitis, conjunctivitis, fungal infection, infection, pneumonia, thrombophlebitis, upper respiratory tract infection, urinary tract infection
Interstitial lung disease	Interstitial lung disease
Metabolism	Blood glucose increased, diabetes mellitus, hypercholesterolaemia, hyperglycaemia, hyperlipidaemia, hypoglycaemia, weight increased
Myopathy	Myopathy, rhabdomyolysis
Neuroleptic malignant syndrome	Neuroleptic malignant syndrome
Neutropenia	Agranulocytosis, febrile neutropenia, granulocytopenia, leukopenia, neutropenia
Peripheral neuropathy	Neuropathy peripheral, paraesthesia
Renal toxicity	Acute kidney injury, azotaemia, oliguria, proteinuria, renal failure, renal impairment
Respiratory failure	Respiratory arrest, respiratory depression, respiratory failure
Seizures	Epilepsy, seizure
Sepsis	Sepsis, septic shock
Serotonin syndrome	Serotonin syndrome
Sleep disturbance	Apnoea, insomnia, sleep disorder
Special senses impairment	Blindness, cataract, deafness, diplopia, dysgeusia, glaucoma, tinnitus, vision blurred, visual acuity reduced, visual field defect, visual impairment
SJS-TEN	Dermatitis bullous, dermatitis exfoliative, drug reaction with eosinophilia and systemic symptoms, erythema multiforme, stevens-johnson syndrome, toxic epidermal necrolysis
Sudden death	Cardiac arrest, sudden death
Suicide	Completed suicide, suicide attempt, suicidal behaviour, suicidal ideation
Thrombotic event vessel unspecified	Cerebral infarction, embolism, thrombosis
Torsade de Pointes	Electrocardiogram QT prolonged, Torsade de Pointes
Venous thrombotic event	Deep vein thrombosis, pulmonary embolism

There were 36 DMEs that were chosen in consultation with medical officers at the US Food and Drug Administration (FDA). Each DME consists of a list of related MedDRA PTs representing a serious adverse event of interest to medical officers at FDA working in postmarket safety.
DME, designated medical events; MedDRA PTs, Medical Dictionary for Regulatory Activities Preferred Terms; SJS-TEN, Stevens-Johnson syndrome-toxic epidermal necrolysis.

drug-target associations, and AEs that are associated with those drugs via FAERS are then mapped to the related targets. In this way, EFFECT can identify targets that are highly associated with a particular AE and vice versa. For the study drugs, all known targets were input into EFFECT, and the software mapped drugs to these targets to search and compute information for comparators at once, outputting a single list of AEs. In the case of multiple targets, PRR and LLR are computed for subset-AE pairs where the subset consists of all case reports in FAERS containing a drug mapped to one of the targets. For example, in the PRR calculation above, the first part ("cases of desvenlafaxine & serotonin syndrome") would be replaced by "cases of subset drugs & serotonin syndrome."

The resulting profile is a list of TAEs coded as MedDRA PTs,[14] each with an associated case count (N), disproportionality score (PRR with 95% confidence interval), and LLR. The MedDRA PTs were then mapped to DMEs using a lookup table. The presence of one MedDRA PT was sufficient to assign the corresponding DME to the TAE profile. DMEs that failed to map via a MedDRA PT were assumed to have zero reports in FAERS and the corresponding N, PRR, and LLR were assigned the value of 0. As PRR has a high type I error rate for small N,[19–21] DMEs with fewer than 30 case reports were also assigned a PRR of 0 to minimize false-positive predictions.[22] In this case, LLR was left unchanged as the adjusted *P* value of 0.05 controls the overall false discovery rate to 0.05 for small N.[18] To compute PRR, LLR, and N across DMEs, statistics were computed at the PT levels, log transformed, aggregated by DME, and means computed.

Search criteria for TAE profiles generated from FAERS reports. For each study drug, EFFECT was queried for all case reports with drugs sharing at least one pharmacological target with the drug and dated prior to marketing approval. MedDRA PTs were then mapped to their respective DMEs, as described above. The specific queries for the 55 study drugs can be found in the **Supplementary Materials**.

Target-adverse event profiles from the FDA drug labels. The same set of comparator drugs used in the FAERS TAE profiles was selected for each study drug based upon shared pharmacological targets. To maintain consistency with the EFFECT software, DrugBank was used as the source of drug-target mappings. For each comparator drug, the most recent FDA product label published prior to the respective study drug approval was identified in DailyMed and text-mined to extract safety content using the I2E software from Linguamatics.[23] Extracted AEs were mapped to DMEs and the proportion of comparator drug labels reporting a DME, referred to as labelscore, was computed for each DME. The DMEs without a MedDRA PT in comparator drug labels were assigned a labelscore of 0.

Current (2017) FDA drug labels were obtained through the National Library of Medicine DailyMed website.[24] Historical labels of original study drugs and comparators were obtained through the DailyMed database of archived labels.[25] Safety content of study drug current labels was extracted as MedDRA PTs using I2E OnDemand software from Linguamatics.[23,26] Safety content of comparator drug historical labels was extracted as MedDRA PTs using the I2E Enterprise software from Linguamatics. Safety content of study drug original labels was extracted manually. Manual extraction was used for historical study drug labels due to the higher error rate of text-mining these labels and the need for direct comparison to the current labels to determine safety label changes. The error rate was deemed acceptable for use on the comparator drug historical labels.

Text-mining query performance analysis. An analysis of the performance of the text-mining query used in Linguamatics to identify adverse drug reactions from the current drug labels was performed to assess the accuracy of the query. Twenty random drugs from the 55 drug set were used to gather baseline performance data for the query. These 20 drugs were manually curated for adverse drug reactions, and the manual curation was compared with the text-mining output. Errors were identified and changes were made to the query to mitigate these errors. This process was repeated three times until the final query used for the study was obtained. A second set of 20 test drugs was additionally evaluated with the final query to obtain final performance statistics.

Target-adverse event profiles from literature reports. The same set of comparator drugs used in the FAERS and label TAE profiles was selected for each study drug based upon shared pharmacological targets. For consistency with the EFFECT software, DrugBank was used as a source of drug-target mappings. For each comparator drug, we queried the EMBASE database from Elsevier[27] to identify comparator drug-associated AEs reported in the literature. EMBASE drug AEs are curated by subject matter experts from full-text literature, abstracts, and conference proceedings. EMBASE curators manually link every AE with the corresponding drug. All terms, including AEs and drugs, are then mapped to Elsevier's controlled terminology, Emtree. Emtree terms that were associated with AEs of interest in this study were manually converted to MedDRA PTs. Using this method, over 3.3 million literature reports were identified for comparator drugs, and from these reports, over 400,000 unique drug-AEs reports were retrieved. These retrieved AEs were mapped to DMEs and the proportion of comparator drugs reported with a DME, referred to as litscore, was computed for each DME. The DMEs that failed to map to a MedDRA PT in the EMBASE query results were assigned a litscore of 0.

All TAE profiles can be found in **Supplementary Materials**. See TA_55_drug_study.zip.

Classification

Constructing the target analysis dataset. Observations consist of 55 drugs with 3 or more years postmarket exposure. Input features (independent variables) are TAE profiles generated for each observation: FAERS case count, FAERS PRR, FAERS LLR, labelscore, and litscore are computed for each DME. Five inputs per drug per DME creates 36*5 = 180 records per drug, which are then pivoted into a single record. The resulting data set has 55 rows and 216 columns (180 inputs and 36 outputs). This allows for exploitation of correlations in the FDA drug labels between DMEs such that data for all DMEs are used as potential features for prediction. For example, historical comparator drug data for the DME "Bone Marrow Failure" can be used to predict the DME "Neutropenia" for a study drug, as "Bone Marrow Failure" and "Neutropenia" frequently co-occur on the FDA drug labels. See **Figure 2c** for DME correlations. The data set and code can be found in the **Supplementary Materials**, file TA_55_drug_study.zip.

Ensemble learning. The above multilabel classification problem (36 DMEs per observation) is transformed to binary relevance (i.e., each DME is treated as an independent classification problem).[28,29]

Initial efforts with a variety of classification and feature selection methods showed highly variable performance with different methods working best for different DMEs (see **Figure S1**). Standard practice is to try various methods, selecting the best performing method under crossvalidation (CV) for final testing on a blinded validation set (external validation). However, two considerations oblige a different approach for this study: (1) the target analysis data set is too small to set aside a dedicated, blinded test set[30]; (2) there are 36 dependent variables instead of the usual one—choosing the best model separately for each output is prone to bias, especially with a small dataset. We therefore use a less biased approach, constructing a voting ensemble of low-complexity classifiers that have a reputation for generalizing well with small data sets and applying this same method to all dependent variables (DMEs).[31–33]

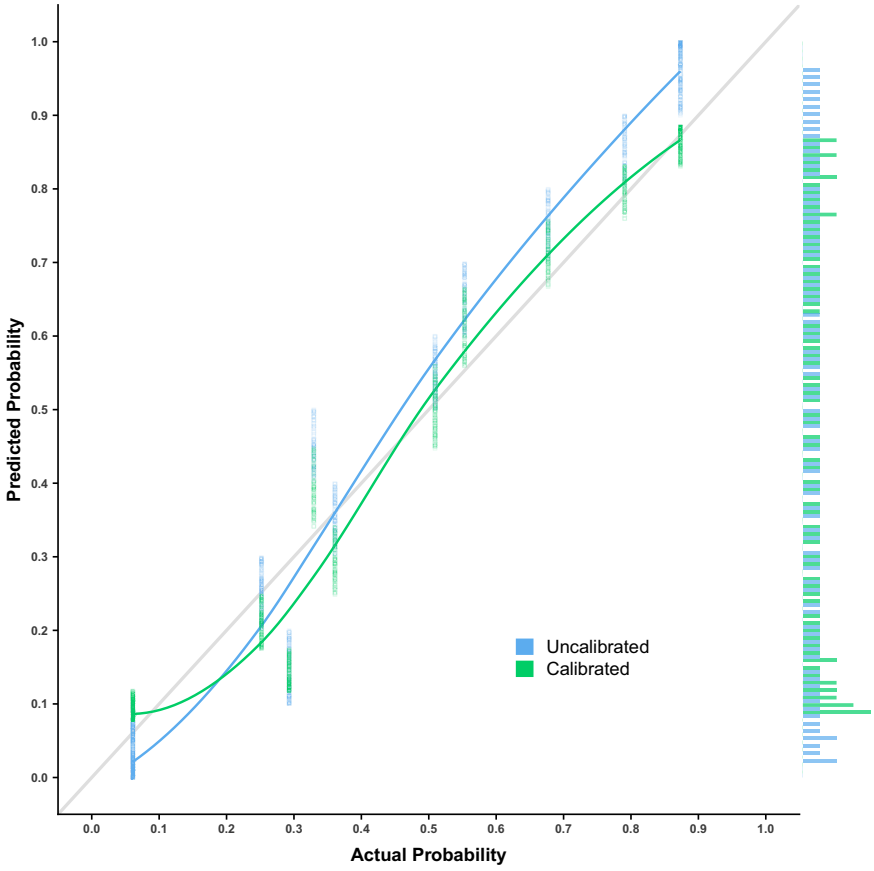

Figure 2 Model calibration. Agreement between predicted probabilities and event probabilities is assessed with calibration plots. Predicted probabilities (dot plot) and loess-smoothed predicted probabilities (line plot) are plotted vs. binned proportion of positive observations (actual probabilities). Predicted probabilities are calibrated with logistic regression. Perfect agreement is indicated by the grey line. A histogram of predicted probabilities is displayed on the right vertical axis.

The classifier has two levels: The first level consists of 4 base models, each trained on the same 10 features most relevant to each DME, selected with regularized regression (described below). The four base models are naive bayes,[34] KNN,[35] SVM with linear kernel,[36] and C4.5.[37] The second level averages the predictions from the first level. For high prevalence DMEs (prevalence of DME > 0.15), each base model was trained under 100 fivefold CVs using the same sampling indices across models. Multiple CVs were used to assess both variance and bias.[38,39] Low-prevalence DMEs were trained under leave-one-out crossvalidation (LOOCV), as the sample size was deemed too small to perform fivefold CV on DMEs with label prevalence < 0.15. The classifier was coded twice in the R programming language[40] using the caret[41] and RWeka[38,42] packages. Classifier performance was similar for the two packages and RWeka was chosen for the final model because of its relative speed and ease of implementation of voting ensembles.

Feature selection. The 180 variable model was reduced to a 10 variable model using elastic-net regularized regression performed with the R package, glmnet.[43] For each DME, regression was performed on 100 boot strapped samples and the 10 most frequent variables selected as features. Two DMEs "Colitis excluding infective" and "Neuroleptic malignant syndrome" with zero and two positive observations, respectively, were ultimately excluded from the final data set because the regression fails with fewer than three observations.

Performance metrics. The target analysis dataset is unbalanced with the prevalence of positives ranging from 0.04 (Colitis excluding infective) to 0.75 (Hypersensitivity) and mean prevalence of 0.34. Additionally, because the FDA product label is a living document such that new AEs are added with postmarket exposure, we do not believe the "negatives" (unlabeled DMEs) in our data set. Indeed, the premise of the study is to use data available at the time of approval to anticipate label changes (postmarket AEs). Therefore, we emphasize performance metrics computed from positive predictions; namely, area under precision recall curve (AUPRC), precision (positive predictive value), recall (sensitivity), and F1 (harmonic mean of precision and recall). Area under receiver operating curve (AUROC), specificity, and Brier score (discussed below) were also computed.

Model calibration. Two methods were used to assess calibration of model predictions (i.e., agreement between binary observations; e.g., presence/absence of a DME on a study drug label) and the predicted probabilities of the observations. (1) Brier score,[44,45] a proper scoring method used in forecasting, was computed for predictions made under LOOCV and fivefold CV. In the binary case, Brier score takes the form of the mean squared error:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=0}^{N}\left(p_i - o_i\right)^2$$

where p_i is the predicted probability that the i^{th} observation belongs to the positive class and o_i is the value of the i^{th} observation encoded in binary with 1 indicating the positive class and 0 the negative class. The range of the Brier score is [0,1] with a lower score indicating superior

calibration. The expected value of the Brier score when $p_i \sim U[0,1]$ is 1/3 and is used as a baseline in this study. (2) Calibration plots were generated for probabilities predicted under LOOCV. Raw and calibrated probabilities were split into 10 bins from 0 to 1 and bin proportion of positive observations ("actual probabilities") plotted vs. loess-smoothed predicted probabilities. Probabilities were calibrated with logistic regression.[46,47]

Agreement between predicted and actual probabilities is assessed using calibration plots (**Figure 2**). Agreement is good for actual probabilities > 0.5, whereas classifier predicted probabilities tend to underpredict actual probabilities when < 0.5.

Safety label changes

For each study drug, the original product label was compared with the current product label and any postmarket changes identified mapped to the DME list. DME label changes were identified among the 55 drugs. Label changes were compared with classifier predictions made under LOOCV and the percentage identified correctly computed. A false-positive error analysis was performed. False-positives were analyzed to determine if they were misclassified as indications, disease symptoms, or disease comorbidities by the review of two physicians. The remaining false positives were reviewed by one investigator (K.B.) to gain insights into causes for misclassification. This analysis included a review of data mining scores in FAERS, review of MedWatch narratives when there was a significant number of cases reported relative to the DME prevalence, and review of drug labels for text-mining errors.

Model code

All model code and data sets are available in **Supplementary Materials**. See TA_55_drug_study.zip.

Software used in this study

TAEs from FAERS reports were generated using the EFFECT from Molecular Health, GmbH.[15] EFFECT aggregates FAERS reports by mapping the active ingredients recorded in each case report to their respective pharmacological targets, as found in DrugBank.[48] The software can then be queried by target or a set of targets to generate a subset of case reports, which can then be used for further analysis.

Text-mining of FDA labels was performed with the I2E software from Linguamatics, version 5.0.[23,26]

AE literature reports were identified using the EMBASE database from Elsevier[27] queried May 2017.

The ensemble model was constructed using the R software, version packages used include C50,[49] car,[41] caret,[41] corrplot,[50] cowplot,[51] DMwR,[52] doParallel,[53] foreach,[54] ggExtra,[55] glmnet,[43] gridExtra,[56] gtable,[57] gtools,[58] Hmisc,[59] kableExtra,[60] kernlab,[61] knitr,[62] MLmetrics,[63] naivebayes,[64] PRROC,[65] RColorBrewer,[66] rJava,[67] rms,[68] RWeka,[42] stringr,[69] tidyverse,[70] and viridis.[71]

RESULTS

Designated medical events

The prevalence of DMEs (the proportion of drugs labeled for a DME) in the dataset is variable, ranging from 0.04 to 0.7

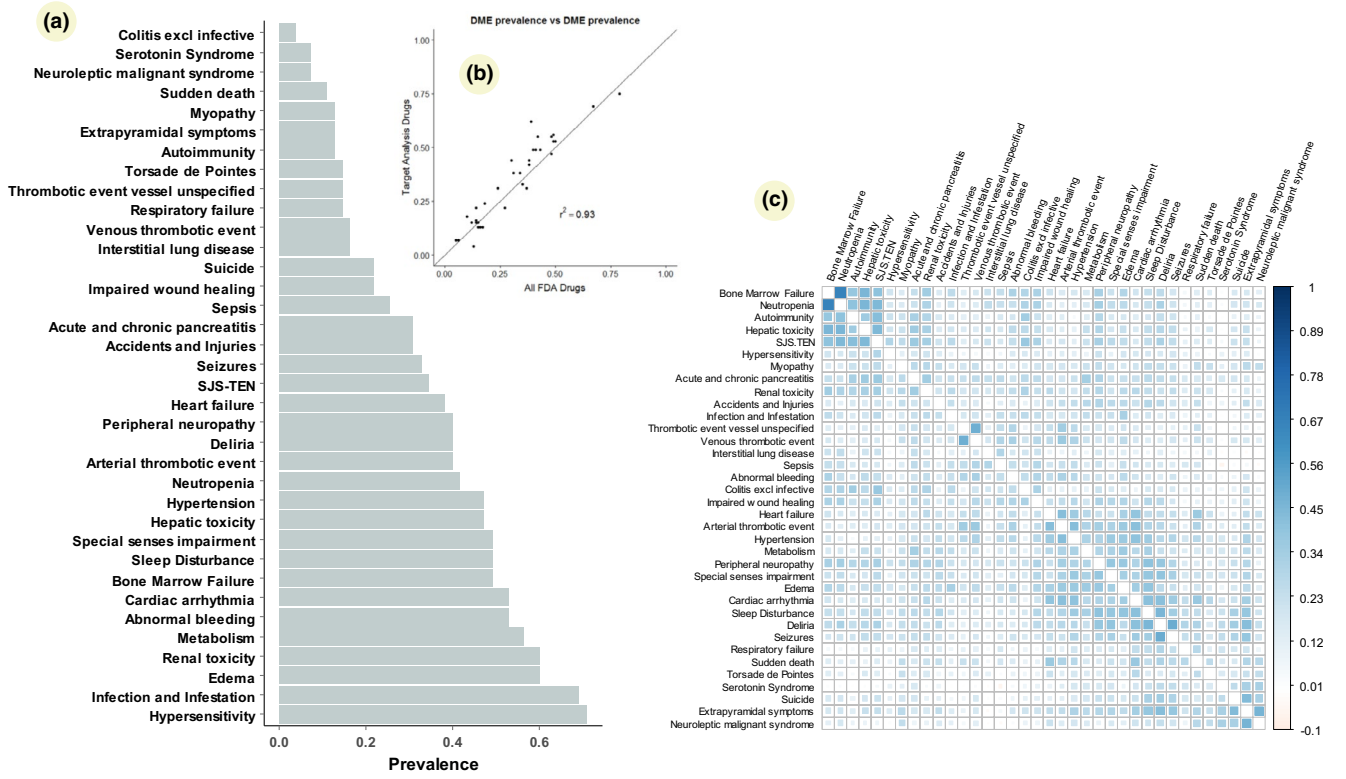

Figure 3 Designated Medical Events. There were 36 Designated Medical Events (DMEs) that were chosen in consultation with medical officers at the US Food and Drug Administration (FDA). Each DME consists of a list of related Medical Dictionary for Regulatory Activities Preferred Terms (MedDRA PTs) representing a serious adverse events (AE) of interest to medical officers and safety evaluators at the FDA working in postmarket safety. See **Supplementary Materials** for the full list of MedDRA PTs comprising the DME list. (**a**) DMEs are plotted in order of prevalence. (**b**) DME prevalence in study drugs is plotted against all FDA drugs.* (**c**) A correlation plot of DMEs for all FDA drugs.**Combination products excluded. SJS-TEN, Stevens-Johnson syndrome- toxic epidermal necrolysis.

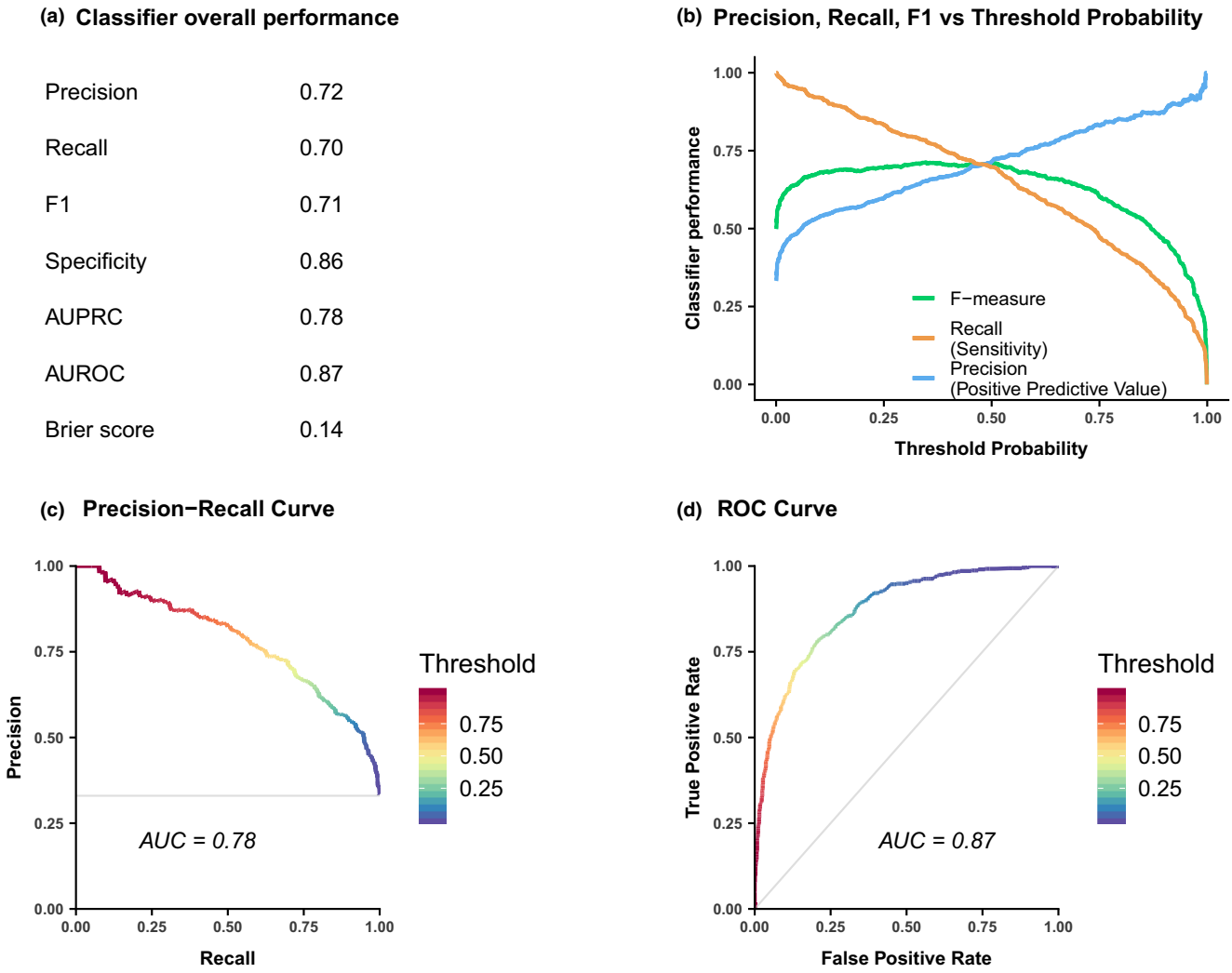

Figure 4 Overall classifier performance. Classifier performance is assessed under leave-one-out crossvalidation (LOOCV) and micro-averaged across designated medical events (DMEs). (**a**) Precision (positive predictive value), recall (sensitivity), F1 (harmonic mean of precision and recall), and specificity are computed using the standard threshold probability of 0.5. Threshold based methods area under precision recall curve (AUPRC), area under receiver operating curve (AUROC), and Brier score are also computed. (**b**) Precision, recall, and F1 are plotted against threshold probability, allowing the user to choose the balance of precision and recall suited to their needs. The threshold maximizing F1 to 0.74 is 0.45. (**c**) Precision is plotted vs. recall. AUPRC = 0.78. The expected value of AUPRC with predicted probabilities drawn from a uniform distribution is the prevalence of the positive class in the data set and serves as baseline (grey line). (**d**) True-positive rate is plotted vs. false-positive rate. AUROC = 0.87. The expected value of AUROC with predicted probabilities drawn from a uniform distribution is 0.5 and serves as baseline (grey line).

(**Figure 3a**). DME prevalences in the set of 55 drugs used in this study are comparable to the full set of FDA approved drugs (**Figure 3b**). **Figure 3c** shows that many DMEs are correlated; for example, Stevens-Johnson syndrome-toxic epidermal necrolysis, a rare drug hypersensitivity affecting skin and mucous membranes, tends to co-occur on drug labels with other hypersensitivities, such as pancreatitis and neutropenia.

Overall classifier performance

Overall performance of the classifier was assessed under LOOCV (**Figure 4**). The classifier shows good overall performance yielding a F1 of 0.71, an AUPRC of 0.78, an AUROC 0.87, and a Brier score of 0.14. Of 1,870 predictions (34 DMEs × 55 drugs) there were 172 false positives, 433 true positives, 188 false negatives, and 1,077 true negatives.

DME-level classifier performance

Classifier performance for each DME was assessed under LOOCV and fivefold CV (**Table S2**, **Figures 4 and 5**). Classifier performance varied with DME. Sixteen DMEs had F1 scores > 0.7 under LOOCV (**Table S2**). Performance tends to improve with DME prevalence; however, some high prevalence (bone marrow failure and special senses impairment) and low prevalence (serotonin syndrome and extrapyramidal symptoms) DMEs performed very well (**Table S2**, **Figure 5**). Within DME variability in classifier performance under fivefold CV tends to decrease with prevalence (**Figure 5**).

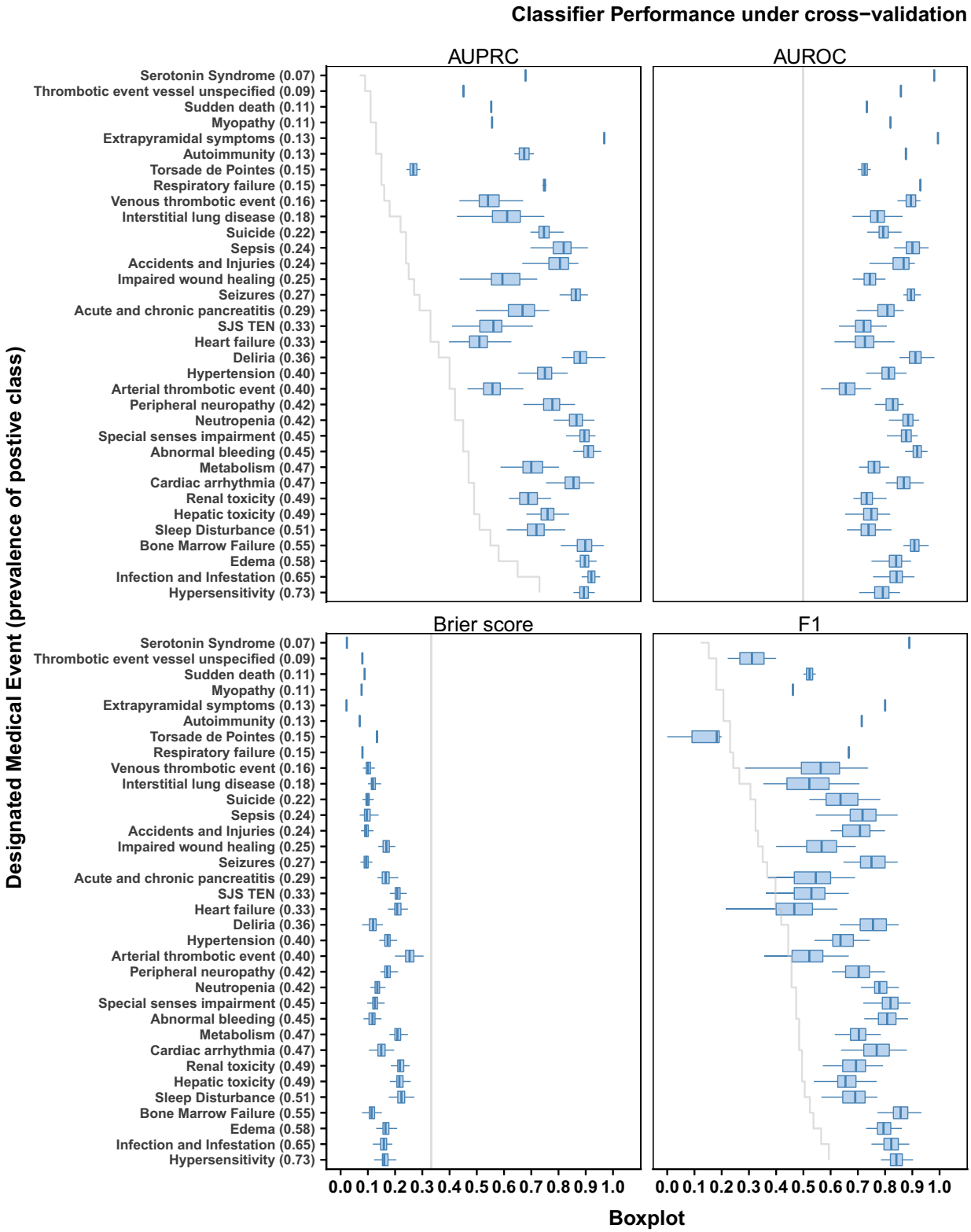

Figure 5 Designated medical event (DME)-level classifier performance. Classifier performance is assessed individually for each DME. High prevalence DMEs (defined as prevalence of positive class > 0.15) are assessed with 100 fivefold crossvalidations. Low prevalence DMEs are assessed with leave-one-out crossvalidation (LOOCV). Mean with one SD is shown for AUPRC, AURPC, Brier score, and F1. The expected value AUPRC, AUROC, Brier score, and F1 when predicted probabilities are drawn from a uniform distribution serves as baseline (grey line) and are, respectively, the prevalence of the positive class, 1/2, 1/3, and prevalence/(0.5 + prevalence). DMEs (vertical axis) are ordered by prevalence of the positive class. AUPRC, area under precision-recall curve; AUROC, area under receiver operating curve. SJS-TEN, Stevens-Johnson syndrome-toxic epidermal necrolysis.

Table 2 Safety label changes

Safety label change	Fraction retrieved (Preds/#SLC)
Abnormal bleeding	2/4
Accidents and injuries	1/3
Acute and chronic pancreatitis	4/6
Arterial thrombotic event	2/3
Autoimmunity	1/2
Bone marrow failure	5/5
Cardiac arrhythmia	2/2
Deliria	6/8
Edema	4/5
Heart failure	1/5
Hepatic toxicity	7/8
Hypersensitivity	12/15
Hypertension	1/1
Impaired wound healing	1/2
Infection and infestation	6/6
Interstitial lung disease	2/6
Metabolism	2/2
Myopathy	1/2
Neutropenia	2/2
Peripheral neuropathy	1/1
Renal toxicity	1/3
Respiratory failure	0/2
Seizures	4/4
Sepsis	1/1
SJS-TEN	6/9
Sleep disturbance	2/4
Special senses impairment	3/3
Sudden death	1/2
Suicide	0/1
Thrombotic event vessel unspecified	0/2
Torsade de Pointes	1/1
Venous thrombotic event	3/3
Total	85/123 (69%)

There were 123 safety label changes that were identified and compared with target analysis predictions under leave-one-out crossvalidation. Eighty-five (69%) were retrieved.
#SLC, count of safety label changes; Preds, number SLCs predicted.
SJS-TEN, Stevens-Johnson syndrome-toxic epidermal necrolysis.

Note that fivefold CV was not performed for low-prevalence (< 0.15) DMEs. Improvement in classifier performance over random predictions and predictions made at the no information rate (the expected performance of the classifier when the DME is predicted naively with a probability of 1) declines with prevalence (**Figure S2**).

Safety label changes

There are 123 safety label changes among the 55 drugs in our dataset (summarized in **Table 2**). Eighty-five (69%) of the safety label changes were retrieved, consistent with the overall sensitivity of the classifier. Additionally, 38 (31%) safety label changes were not retrieved. See **Table S3** for the full list of safety label changes and classifier predictions.

False positive analysis

The classifier made 172 false-positive predictions under LOOCV. Three of these (1.7%) were determined to be indication related. Thirty-five (20.2%) were determined to be symptoms of the disease indication or disease comorbidities. Twenty-seven (15.6%) were labeled or captured by another preferred term not listed in the DME category. Thirty-four cases (19.7%) were considered new signals that warrant further investigation in the opinion of the physician reviewer (K.B.) based on a high case count, disproportionality score, a review of case narratives, and mechanistic plausibility. Of the remaining 74 false positives, 9 (5.2%) resulted from the selection of comparators that had indications and comorbidities unrelated to the drug of interest's indications.

Text-mining query performance analysis

A formal analysis of the text-mining query was performed (see **Figure S3**). After 3 rounds of improvement, the final query used for this study had a sensitivity of 0.98, a precision of 0.94, and an F1 score of 0.96 when tested on 20 random drugs from this study used to train the query. When tested on 20 different random drugs from this study, the final query had a sensitivity of 0.91, a precision of 0.90, and an F1 score of 0.90. The most common errors included identification of a comorbidity as an AE (27%), identification of section headers (i.e., "Gastrointestinal Disorders") as AEs (21%), and inaccurate mapping of terms from free text to MedDRA PTs (16%; **Figure S3b**). This query is currently available in the Linguamatics OnDemand software for FDA Product Label querying, and improvements are ongoing.

DISCUSSION

The creation of TAE profiles by using comparator drugs (not necessarily in the same class) that share similar molecular target activity shows promise in identifying adverse drug events (adverse drug reactions) for augmented pharmacovigilance. The model has the potential to expedite safety communication and improve public health via early identification of postmarket safety concerns. Additionally, the model identified future safety label changes and new unlabeled safety signals that warrant further review. The model has the potential to add efficiency for safety evaluators by automating target and class analyses summarizing FAERS data mining results, label comparisons, and literature reports.

Model performance varied with DME, and DME prevalence was not a key predictor of performance. Serotonin syndrome represents a DME where the mechanisms (receptors and actions) that enhance serotonin neurotransmission and precipitate the syndrome are well-understood.[72] Other DMEs, such as idiosyncratic drug-induced pancreatitis, have multiple mechanisms (e.g., hypertriglyceridemia, duct obstruction, and autoimmunity) that can cause pancreatitis. The key molecular targets involved in the development of pancreatitis have not been clearly identified. Therefore, a cohort of drugs linked to pancreatitis has more variability, and thus the DME pancreatitis performed less accurately than the serotonin syndrome DME and other more target-specific DMEs.

Additionally, we significantly reduced the number of PTs used as compared with our previous study. For studies such as this one, having a representative set of PTs is critical. However, having too many PTs may contribute unnecessarily to the false-positive rate, as seen in our previous study and additional unpublished analysis. Limiting our study to 167 PTs may lead to some missed signals and predictions. However, after extensive conversations with medical officers and analysis of labeled and spontaneously reported AEs, this study eliminated many terms that were ambiguous or not clearly related to the DME. Future work could evaluate the DME list and related PTs more extensively to insure the captured terms are truly representative of the DME mechanisms and the data sources.

The selection of comparators is a key step in the process. Some drugs that performed poorly (had many false-positive predictions) may have performed better if a different set of comparators was chosen. For example, a large amount of comparators may include drugs that are not similar to the drug of interest, and therefore AEs that are not relevant may be predicted. An algorithm that selects comparators from the most similar 3–5 drugs and/or all existing class members warrants further study.

One of the main factors responsible for classification errors was text-mining error. An analysis of the text-mining query used to extract AE data from the FDA labels found that the query falsely identifies indication-related inclusion criteria from the clinical trial descriptions as AEs. Patient characteristics, comorbidities, and disease-related symptoms are also captured as AEs for the disease indication, sometimes further complicated by disease indications of comparators. Additionally, some AE terms were not properly mapped to MedDRA terms, and therefore DMEs. Further enhancement of the text-mining query to better capture patient characteristics negate drug indications, and properly map AEs to MedDRA terms will improve performance, as improving the quality of the input data may lead to more accurate predictions.

Finally, additional features related to target activity or structural similarity to the drug of interest may be considered for inclusion to improve model accuracy. These features, such as Tanimoto coefficients or shared key substructures, could be used in comparator selection; similarly, shared pathway signaling, or gene expression could assist comparator selection. Alternatively, chemical features,[7,8] gene expression,[8] or pathway signaling[8] could be used as additional features for machine learning. Selection of comparators that share structural features may have activity at similar targets. Via this theory, idiosyncratic reactions, such as pancreatitis, may be more accurately identified as shared AEs that result from shared unknown targets. Last, after evaluating modifications to the algorithm, we look to evaluate the addition of features from other databases, including the electronic health record,[9] prescription records,[10] and social media.[11]

CONCLUSIONS

The use of TAE profiles by the selection of comparator drugs demonstrates promise as a method to identify AEs unknown at the time of product approval. Augmented pharmacovigilance tools, such as this one, can save time and resources by identifying potential postmarketing safety concerns. This model used an ensemble machine learning applied to data from drug labels, literature, and FAERS. Further refinement to improve accuracy could evaluate AE selection, comparator selection, label text-mining, and literature queries.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

The authors would like to acknowledge the following contributions: FDA Office of Surveillance and Epidemiology for help developing the designated medical event list; Darrell Abernethy, Ram Tiwari, Ted Guo, Scott Proestel, and Joseph Tonning of the FDA, and Eibe Frank of the University of Waikato for helpful discussions.

FUNDING

P.S., R.R., D.G.S., and K.B. were funded by the US Food and Drug Administration. D.B.J. and T.G.S. were funded by the Molecular Health, GmbH.

CONFLICTS OF INTEREST

D.B.J. is an employee of Molecular Health, GmbH, a shareholder in Molecular Health, GmbH, and inventor of the EFFECT technology; T.G.S. is an employee of Molecular Health, GmbH and a shareholder in Molecular Health, GmbH and inventor of the EFFECT technology. All other authors declare no competing interests for this work.

AUTHOR CONTRIBUTIONS

P.S., R.R., R.L., D.G.S., and K.B. wrote the manuscript. P.S., R.R., K.B., and R.L. designed the research. P.S. and R.R performed the research. P.S., R.R., D.G.S., and K.B. analyzed the data. D.B.J. and T.G.S. contributed new reagents/analytical tools.

DISCLAIMERS

This study reflects the views of the authors and should not be construed to represent the views or policies of the FDA. As Associate Editor of *Clinical Pharmacology and Therapeutics*, David G. Strauss was not involved in the review or decision process for this paper.

1. Downing, N.S. *et al*. Postmarket safety events among novel therapeutics approved by the US food and drug administration between 2001 and 2010. *JAMA* **317**, 1854–1863 (2017).
2. Jiang, G., Wang, L., Liu, H., Solbrig, H.R. & Chute, C.G.J.M. Building a knowledge base of severe adverse drug events based on AERS reporting data using semantic web technologies. *Stud. Health Technol. Inform.* **2013**, 496–500 (2013).
3. Xu, R. & Wang, Q. Automatic signal extraction, prioritizing and filtering approaches in detecting post-marketing cardiovascular events associated with targeted cancer drugs from the FDA Adverse Event Reporting System (FAERS). *J. Biomed. Inform.* **47**, 171–177 (2014).
4. Schotland, P. *et al*. Target-adverse event profiles to augment pharmacovigilance: a pilot study with six new molecular entities. *CPT Pharm. Syst. Pharmacol.* **7**, 809–817 (2018).

5. Soldatos, T.G., Taglang, G. & Jackson, D.B. In silico profiling of clinical phenotypes for human targets using adverse event data. *High Through.* **7**, 37, (2018).
6. Shang, N., Xu, H., Rindflesch, T.C. & Cohen, T. Identifying plausible adverse drug reactions using knowledge extracted from the literature. *J. Biomed. Inform.* **52**, 293–310 (2014).
7. Liu, M. *et al*. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J. Am. Med. Inform. Assoc.* **19**, e28–e35 (2012).
8. Wang, Z., Clark, N.R. & Ma'ayan, A. Drug-induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* **32**, 2338–2345 (2016).
9. Wang, L., Rastegar-Mojarad, M., Liu, S., Zhang, H. & Hongfang, L. Discovering adverse drug events combining spontaneous reports with electronic medical records: a case study of conventional DMARDs and biologics for rheumatoid arthritis. *AMIA Jt. Summits Transl. Sci. Proc.* **2017**, 95–103 (2017).
10. Zhan, C., Roughead, E., Liu, L., Pratt, N. & Li, J. A data-driven method to detect adverse drug events from prescription data. *J. Biomed. Inform.* **85**, 10–20 (2018).
11. Liu, J., Zhao, S. & Zhao, X. An ensemble method for extracting adverse drug events from social media. *Artif. Intell. Med.* **70**, 62–76 (2016).
12. Vanderwall, D.E. *et al*. Molecular clinical safety intelligence: a system for bridging clinically focused safety knowledge to early-stage drug discovery–the GSK experience. *Drug Discov. Today* **16**, 646–653 (2011).
13. US Food and Drug Administration (FDA). *FDA Structured Product Labeling* <https://www.fda.gov/forindustry/datastandards/structuredproductlabeling/> (FDA, Silver Spring, MD, 2018).
14. International Conference on Harmonisation (ICH). *MedDRA: Medical Dictionary for Regulatory Activities* <http://www.meddra.org/> (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Silver Spring, MD, 2017).
15. Molecular Health, I. *Molecular Health MH EFFECT. Vol. 2016* <https://www.molecularhealth.com/us/applications-and-solutions/mh-effect/> (Molecular Health, I, Heidelberg, Germany, 2016).
16. van Puijenbroek, E.P. *et al*. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* **11**, 3–10 (2002).
17. Evans, S.J., Waller, P.C. & Davis, S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol. Drug Saf.* **10**, 483–486 (2001).
18. Huang, L., Zalkikar, J. & Tiwari, R.C. Likelihood ratio test-based method for signal detection in drug classes using FDA's AERS database. *J. Biopharma. Statist.* **23**, 178–200 (2013).
19. Almenoff, J.S., LaCroix, K.K., Yuen, N.A., Fram, D. & DuMouchel, W. Comparative performance of two quantitative safety signalling methods: implications for use in a pharmacovigilance department. *Drug Saf.* **29**, 875–887 (2006).
20. Harpaz, R. *et al*. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin. Pharmacol. Ther.* **93**, 539–546 (2013).
21. Dumouchel, W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am. Stat.* **53**, 177–190 (1999).
22. US Food and Drug Administration (FDA). Guidance for Industry: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment. (ed. Services, U.H.a.H.) (2005).
23. Linguamatics Ltd. *Linguamatics I2E Natural Language Processing Software*. (Lingumatics, Cambridge, UK, 2018).
24. National Library of Medicine. *DailyMed SPL Resources: Download All Drug Labels* <https://dailymed.nlm.nih.gov/dailymed/spl-resources-all-drug-labels.cfm> (National Library of Medicine, Bethesda, MD, 2018).
25. National Library of Medicine. *DailyMed Label Archives* <https://dailymed.nlm.nih.gov/dailymed/archives/index.cfm> (National Library of Medicine, Bethesda, MD, 2018).
26. Milward, D. *et al*. Ontology-based interactive information extraction from scientific abstracts. *Int. J. Genomics* **6**, 67–71 (2005).
27. Elsevier. *EMBASE* (Elsevier, Oxford, UK, 2018).
28. Godbole, S. & Sarawagi, S. Discriminative methods for multi-labeled classification. in Pacific-Asia conference on knowledge discovery and data mining 22–30 (Springer, New York, NY, 2004).
29. Tsoumakas, G. & Katakis, I. Multi-label classification: an overview. *Int. J. Data Warehousing Mining (IJDWM)* **3**, 1–13 (2007).
30. Steyerberg, E.W. Validation in prediction research: the waste by data-splitting. *J. Clin. Epidemiol.* **103**, 131–133 (2018).
31. Zhou, Z.-H. *Ensemble Methods: Foundations and Algorithms* (Chapman and Hall/CRC, New York, NY, 2012). ISBN: 1439830053.
32. Kittler, J., Hatef, M., Duin, R.P. & Matas, J. On combining classifiers. *IEEE Transact. Patt. Analysis Mach. Intel.* **20**, 226–239 (1998).
33. Kuncheva, L.I. *Combining pattern classifiers: methods and algorithms* (John Wiley & Sons, Hoboken, NJ, 2004). ISBN: 0-471-21078-1.
34. John, G.H. & Langley, P. Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence* 338-345 (Morgan Kaufmann Publishers Inc., Burlington, MA, 1995). ISBN: 1558603859.
35. Aha, D.W., Kibler, D. & Albert, M.K. Instance-based learning algorithms. *Machine Learn.* **6**, 37–66 (1991).
36. Platt, J.C. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods - Support Vector Learning*, (eds. Scholkopf, B., Burges, C. & Smola, A.) 185–208 (MIT press, Cambridge, MA, 1999).
37. Quinlan, J.R. C4. 5: Programming for machine learning. *Morgan Kauffmann* **38**, 48 (1993).
38. Witten, I.H. & Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques With Java Implementations* (Morgan Kaufmann Publishers, San Francisco, CA, 2000). ISBN: 9781558605527.
39. Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning* (Springer series in statistics, New York, NY, 2001).
40. Team, R.C. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2018).
41. Wing, M.K.C.F.J. *et al*. caret: Classification and Regression Training <https://cran.r-project.org/web/packages/caret/> (2018).
42. Hornik, K. RWeka: R/Weka Interface. R package version 0.4-38 <https://cran.r-project.org/web/packages/RWeka/> (2018).
43. Friedman, J. *et al*. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. R package version 2.0-16 <https://cran.r-project.org/web/packages/glmnet/index.html> (2018).
44. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **78**, 1–3 (1950).
45. Rufibach, K. Use of Brier score to assess binary predictions. *J. Clin. Epidemiol.* **63**, 938–939 (2010).
46. Harrell, F.E., Lee, K.L. & Mark, D.B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **15**, 361–387 (1996).
47. Spiegelhalter, D.J. Probabilistic prediction in patient management and clinical trials. *Statist. Med.* **5**, 421–433 (1986).
48. Wishart, D.S. *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
49. Kuhn, M. & Quinlan, R. C50: C5.0 Decision Trees and Rule-Based Models. R package version 0.1.2 <https://cran.r-project.org/web/packages/C50/> (2018).
50. Wei, T. & Simko, V. R package "corrplot": Visualization of a Correlation. Matrix. R package (Version 0.84) <https://cran.r-project.org/web/packages/corrplot/index.html> (2017).
51. Wilke, C.O. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 0.9.2 <https://cran.r-project.org/web/packages/cowplot/index.html> (2017).

52. Torgo, L. & Torgo, M.L. Package 'DMwR'. Comprehensive R Archive Network. R package <https://cran.r-project.org/web/packages/DMwR/> (2013).
53. Corporation, M. & Weston, S. doParallel: Foreach Parallel Adaptor for the 'parallel'. Package. R package version 1.0.11 <https://cran.r-project.org/web/packages/doParallel/> (2017).
54. Microsoft & Weston. S. foreach: Provides Foreach Looping Construct for R (2017).
55. Attali, D.B. C. ggExtra: Add marginal histograms to 'ggplot2', and more 'ggplot2'enhancements. R package version 0.8 <https://CRAN.R-project.org/package=ggExtra> (2018).
56. Auguie, B. gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3 <https://cran.r-project.org/web/packages/gridExtra/index.html> (2017).
57. Wickham, H. gtable: Arrange 'Grobs' in Tables. R package version 0.2.0 <https://cran.r-project.org/web/packages/gtable/index.html> (2016).
58. Warnes, G.R., Bolker, B. & Lumley, T. gtools: Various R Programming Tools. R package version 3.5.0 <https://cran.r-project.org/web/packages/gtools/index.html> (2015).
59. Harrell Jr, F. *et al*. Hmisc: Harrell Miscellaneous. R package version 4.1-1 <https://cran.r-project.org/web/packages/Hmisc/index.html> (2018).
60. Zhu, H. kableExtra: Construct Complex Table with 'kable' and Pipe. Syntax. R package version 0.9.0 <https://cran.r-project.org/web/packages/kableExtra/index.html> (2018).
61. Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. kernlab-an S4 package for kernel methods in R. *J. Stat. Soft.* **11**, 1–20 (2004).
62. Xie, Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.20 <https://cran.r-project.org/web/packages/knitr/index.html> (2018).
63. Yan, Y. MLmetrics: Machine learning evaluation metrics. R package version 1.1.1 <https://cran.r-project.org/web/packages/MLmetrics/> (2016).
64. Majka, M. naivebayes: High Performance Implementation of the Naive Bayes Algorithm. R package version 0.9.2 <https://cran.r-project.org/web/packages/naivebayes/> (2018).
65. Grau, J.& Keilwagen, J. PRROC: Precision-Recall and ROC Curves for Weighted and Unweighted Data. R package version 1.3.1 <https://cran.r-project.org/web/packages/PRROC/> (2018).
66. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. R package version 1.1-2 <https://cran.r-project.org/web/packages/RColorBrewer/index.html> (2014).
67. Urbanek, S. rJava: Low-level R to Java interface. R package version 0.9-9 <https://CRAN.R-project.org/package=rJava> (2017).
68. Jr, F.E.H. rms: Regression Modeling, Strategies. R package version 5.1-2 <https://CRAN.R-project.org/package=rms> (2018).
69. Wickham, H. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.1 <https://cran.r-project.org/web/packages/stringr/index.html> (2018).
70. Wickham, H. tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1 <https://cran.r-project.org/web/packages/tidyverse/index.html> (2017).
71. Garnier, S. viridis: Default Color Maps from 'matplotlib'. R package Version 0.4. 0 <https://cran.r-project.org/web/packages/viridis/index.html> (2017).
72. Racz, R., Soldatos, T.G., Jackson, D. & Burkhart, K. Association between serotonin syndrome and second-generation antipsychotics via pharmacological target-adverse event analysis. *Clin. Transl. Sci.* **11**, 322–329 (2018).