



Towards Automating Adverse Event Review: A Prediction Model for Case Report Utility

Monica A. Muñoz^{1,2} · Gerald J. Dal Pan¹ · Yu-Jung Jenny Wei^{2,4} · Chris Delcher³ · Hong Xiao² · Cindy M. Kortepeter¹ · Almut G. Winterstein^{2,4,5}

Published online: 7 January 2020

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

Abstract

Introduction The rapidly expanding size of the Food and Drug Administration's (FDA) Adverse Event Reporting System database requires modernized pharmacovigilance practices. Techniques to systematically identify high utility individual case safety reports (ICSRs) will support safety signal management.

Objectives The aim of this study was to develop and validate a model predictive of an ICSR's pharmacovigilance utility (PVU).

Methods PVU was operationalized as an ICSR's inclusion in an FDA-authored pharmacovigilance review's case series supporting a recommendation to modify product labeling. Multivariable logistic regression models were used to examine the association between PVU and ICSR features. The best performing model was selected for bootstrapping validation. As a sensitivity analysis, we evaluated the model's performance across subgroups of safety issues.

Results We identified 10,381 ICSRs evaluated in 69 pharmacovigilance reviews, of which 2115 ICSRs were included in a case series. The strongest predictors of ICSR inclusion were reporting of a designated medical event (odds ratio (OR) 1.93, 95% CI 1.54–2.43) and positive dechallenge (OR 1.67, 95% CI 1.50–1.87). The strongest predictors of ICSR exclusion were death reported as the only outcome (OR 2.72, 95% CI 1.76–4.35), more than three suspect products (OR 2.69, 95% CI 2.23–3.24), and > 15 preferred terms reported (OR 2.69, 95% CI 1.90–3.82). The validated model showed modest discriminative ability (C-statistic of 0.71). Our sensitivity analysis demonstrated heterogeneity in model performance by safety issue (C-statistic range 0.58–0.74).

Conclusions Our model demonstrated the feasibility of developing a tool predictive of ICSR utility. The model's modest discriminative ability highlights opportunities for further enhancement and suggests algorithms tailored to safety issues may be beneficial.

1 Introduction

For nearly 50 years, the US Food and Drug Administration's (FDA) analysis of individual case safety reports (ICSRs) has been an integral component of postmarketing drug surveillance [1, 2]. The rapidly expanding size of the FDA Adverse Event Reporting System (FAERS) database, the repository for these reports, and the complexity of information contained in each report call for automated algorithms for

The views expressed are those of the authors and do not necessarily represent the position of, nor imply endorsement from, the US Food and Drug Administration or the US Government.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40264-019-00897-0>) contains supplementary material, which is available to authorized users.

✉ Monica A. Muñoz
monica.munoz@fda.hhs.gov

¹ Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

² Department of Pharmaceutical Outcomes and Policy, College of Pharmacy, University of Florida, Gainesville, FL, USA

³ Department of Pharmacy Practice and Science, College of Pharmacy, University of Kentucky, Lexington, KY, USA

⁴ Center for Drug Evaluation and Safety (CoDES), University of Florida, Gainesville, FL, USA

⁵ Department of Epidemiology, College of Medicine and College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA

Key Points

Techniques to systematically identify and distinguish higher utility individual case safety reports (ICSRs) from lower utility ones will improve timeliness in the management of safety signals.

We developed a predictive model to prioritize ICSRs for manual review that performed better than using data field completeness alone.

The model's modest performance demonstrated the feasibility of this approach, but also highlighted opportunities for further refinement such as extracting additional predictors from unstructured data and developing algorithms for particular safety issues.

effective identification and evaluation of safety signals. The volume of ICSRs in FAERS precludes a manual review of all ICSRs. In 2018 alone, over 2 million ICSRs were received in FAERS [3]. While automated approaches based on disproportionality analyses are routinely utilized to identify patterns of potential interest in FAERS [4], the practices used to prioritize case-by-case review of ICSRs when determining if signals identified from disproportionality analyses warrant further evaluation rely primarily on individual experience or application of completeness metrics such as *vigiGrade* [5]. This prioritization can include considerations such as severity of outcome (e.g., death) and patient characteristics (e.g., pediatrics). The intent of prioritization is to identify ICSRs with the highest 'pharmacovigilance utility'—those ICSRs with substantial implications on public health—as early as possible. The identification of these ICSRs earlier, particularly when there are hundreds or thousands of ICSRs underlying a signal, can allow for reprioritization of resources commensurate with the anticipated safety implications, and consequently, timely regulatory actions.

Safety signals generated from FAERS are evaluated through a careful review of the ICSRs by specialized safety staff. After a signal has been identified, safety experts search FAERS to identify potential ICSRs for inclusion in a case series that can inform that safety signal. Reviewers then assess for duplicate ICSRs (i.e., ICSRs reported by more than one reporter that describe the same suspect product, event, and patient), apply a case definition specific to the safety issue being evaluated (e.g., do the case details support a diagnosis of say, anaphylaxis?), and assess the causal relationship between the suspected product and the adverse event [6]. The resulting case series is then evaluated by the safety team in the context of other relevant data streams (e.g., preclinical data, clinical trial data, drug utilization) and current product labeling. Findings from the safety analyses are documented and include recommendations for regulatory

action, if warranted. Changes to the product's approved professional label, sometimes referred to as the package insert, are the most common regulatory actions recommended in FDA's pharmacovigilance reviews. These recommendations are guided by several safety guidance documents that provide considerations regarding placement, content, and format of adverse reaction information in labeling and Medication Guides [7].

Defining the construct of 'pharmacovigilance utility' (PVU) at the level of an ICSR is generally absent from the literature. Previous work has successfully developed measures to quantify the completeness of clinical documentation in ICSRs, but completeness is only one contributing factor in a more holistic concept of utility that considers the potential for information to offer true insights into drug safety [5, 8, 9]. Retrospective data on whether an ICSR contributed to a regulatory action (e.g., labeling change, product recall, initiation of further studies) offers a post-hoc method for understanding the factors associated with PVU. Using this conceptualization of PVU, we compiled a training set of FAERS ICSRs from the FDA's prior pharmacovigilance reviews. Thus, the objective of this study was to develop and validate a model predictive of an ICSR's PVU.

2 Methods

2.1 Data Source: Pharmacovigilance Reviews and FDA Adverse Event Reporting System (FAERS) Individual Case Safety Reports (ICSRs)

All pharmacovigilance reviews completed from January 1, 2016, to December 31, 2016, were extracted from FDA's internal document repository. FDA's pharmacovigilance reviews contain the pertinent regulatory background, FAERS search criteria (if applicable), search results, case definition, causality assessment criteria and results, the resulting case series, and regulatory recommendations (e.g., labeling revision, continue routine pharmacovigilance). Each pharmacovigilance review was assessed for the following exclusion criteria: (1) no regulatory action recommended beyond continuation of routine pharmacovigilance, (2) FAERS data was not the data source evaluated (e.g., the data source was a sponsor submission), (3) a causality assessment of all ICSRs retrieved from the FAERS query was not performed (i.e., the review documented simple counts or a sample of ICSRs were reviewed), and (4) a review was not focused on a specific safety issue (e.g., review of all pediatric adverse events for a particular product). Reviews that did not include a causality assessment were excluded because there is no distinction between cases included or excluded from the case series. These types of reviews tend to provide simple ICSR counts without regulatory recommendations (exclusion

criteria 1), rather than in-depth evaluations of safety issues supporting regulatory actions.

After identifying reviews eligible for inclusion, we used the FAERS search criteria in the review to identify and extract all relevant ICSRs that were available for evaluation by the reviewer. The most recent version of an ICSR was used for model development. This replication was necessary because reviews documented the number of ICSRs retrieved by the original reviewer's search and the unique FAERS case number of the ICSRs included in the case series but did not document the FAERS case numbers of ICSRs that were excluded from the case series. The number of ICSRs retrieved in the replicated searches was compared with the original review's search results. Deviations $> 5\%$ in the number of ICSRs retrieved when comparing the de novo searches with results documented in the review were evaluated for a root cause (e.g., follow-up information resulted in ICSRs no longer meeting the search criteria). Reviews with unreconciled deviations were excluded.

2.2 Dependent Variable: Pharmacovigilance Utility (PVU)

PVU was operationalized as an ICSR's inclusion in an FDA-authored pharmacovigilance review's case series supporting a recommendation to modify product labeling. ICSRs included in the pharmacovigilance reviews identified in 2.1 were pooled across reviews and classified as included (PVU = 1) and excluded ICSRs (PVU = 0) for model development.

2.3 Independent Variables: Determinants of PVU

The potential determinants of PVU included aspects of ICSR completeness (i.e., elements of a good ICSR as defined by FDA guidance [6]), features identified from the literature, and variables identified by expert opinion. We used separate variables for measures of completeness rather than a single score, such as the *vigiGrade* completeness score [5], to allow certain elements to have greater importance. We also considered variables evaluated in a previously developed model predictive of a causality assessment of at least 'possible' as measured on a modified WHO-UMC scale in a prior evaluation of FAERS data [10]. The additional determinants we considered based on expert opinion included variants of serious outcomes¹ (e.g., any serious outcome, death as the only reported outcome), reporter country (USA vs other country), ICSR type² (Direct, Expedited, Non-Expedited),

¹ Serious outcomes are defined by US regulations (CFRs 314.80, 600.80) and include death, life-threatening, hospitalization, disability, congenital anomaly, and other serious outcomes.

² Expedited and non-expedited ICSRs are reports that pharmaceutical manufacturers are required to submit by regulation (CFRs 314.80, 600.80).

positive rechallenge, positive dechallenge, literature article, more than one follow-up ICSR, and presence of designated medical events (DMEs) preferred terms. DMEs are adverse events that are considered serious and may often be caused by exposure to drugs. DMEs are used by pharmacovigilance experts to help focus attention on important events and prioritize pharmacovigilance activities (e.g., signal detection). Examples include Stevens–Johnson syndrome, acute hepatic failure, and torsades de pointes [11]. Most variables were derived from structured data fields, but the free-text narrative was also used for operationalizing three variables (literature article, limited narrative, and curated narrative terms). All variables evaluated are described in Supplementary Materials Table 1 and 2 [see electronic supplementary material (ESM)].

2.4 Data Analysis

We characterized pharmacovigilance reviews meeting the inclusion criteria by the following: product(s) reviewed, safety issue evaluated, number of ICSRs retrieved by the search, and number of ICSRs included in the case series. The highest level of labeling recommended was also collected, with Boxed Warning $>$ Warning and Precautions $>$ Contraindication $>$ Adverse Reactions $>$ other sections as the hierarchy [12]. We then compared patient, suspect product, adverse event, and reporter characteristics for differences between ICSRs classified as having PVU versus those without.

Univariate analyses were used to examine crude associations between the potential determinants and PVU. Variables demonstrating limited face validity or no predictive value in the univariate analyses were excluded from multivariable analyses. Three models were considered: a model including completeness variables only, a full model including all available predictors, and a parsimonious model developed via fast backward technique that retained all variables with a p value < 0.15 . We reported odds ratios (ORs) and associated 95% confidence intervals (CIs). We evaluated each model's ability to discriminate between cases selected for inclusion in a case series by using the receiver operating curve (C-statistic), calibration using the Hosmer–Lemeshow test, and the Akaike information criterion (AIC). After evaluating the performance of the three models, the best model was selected for validation. To correct for optimism resulting from using the same data for development and validation, we refitted the final model in 100 bootstrap samples of the original data set [13, 14]. The resulting model parameters were then reapplied to the original dataset, which yielded the optimism-adjusted C-statistic. Statistical analyses were conducted using SAS[®] 9.4 (SAS Institute Inc., Cary, NC, USA).

We conducted two sensitivity analyses to evaluate the robustness of the validated model. In the first analysis, we

evaluated the potential impact of correlated data (i.e., a safety reviewer's decision to include an ICSR in a case series within a review may be correlated) using a generalized estimating equation (GEE) model that considered each review as a cluster. The resulting parameter estimates and overall area under the curve (AUC) were generated for comparison to the validated model. In the second sensitivity analysis, we evaluated for heterogeneity in the model's discrimination by subgroups of safety issues. The safety issues selected for further evaluation were the most common topics of evaluation in the review pool. ICSRs from reviews evaluating hypersensitivity/anaphylaxis, drug-induced liver injury, cardiovascular events (including ventricular arrhythmias, torsades de pointes, QT prolongation, hypotension, AV block, heart failure), and events without acute life-threatening outcomes (including alopecia, application-site pigmentation changes/scarring, false-positive drug screen, fingerprint loss, weight gain) were selected for this analysis.

3 Results

3.1 Pharmacovigilance Reviews and ICSR Pool

Of the 311 reviews documented in the study period, we identified 69 reviews that met the review inclusion criteria. These reviews contained a total of 10,381 ICSRs. The primary reasons for review exclusions included no regulatory recommendations ($n=106$), reviews evaluated did not include FAERS data, but evaluated another data source ($n=29$), review was not focused on a specific safety issue (e.g., reviews of all pediatric adverse events) ($n=36$), the ICSR search was not able to be precisely replicated ($n=27$), not all ICSRs retrieved by the search were manually evaluated in the review ($n=23$), and duplicate review documents ($n=21$).

Characteristics of the 69 included reviews are provided in Table 1. The median number of ICSRs reviewed and included in a series across reviews was 89.5 and 16.5, respectively. The proportion of ICSRs included in a review (ICSRs included in a series/all ICSRs retrieved by the search) ranged from 1.5 to 90.7%, with a median of 22.2%. The median number of duplicate ICSRs identified in a review was eight ICSRs, ranging from zero to 190. The most frequent therapeutic areas of reviewed products included cardiology ($n=11$), oncology/hematology ($n=11$), endocrine/metabolic ($n=9$), gastrointestinal ($n=8$), and neurology ($n=8$). Most reviews evaluated a single product (55/69, 80%), 12 evaluated a class of products, and two reviews evaluated drug–drug interactions. Eight reviews evaluated therapeutic biologic products. The most frequently reviewed safety issues classified by MedDRA® System Organ Class (SOC) included 'skin and subcutaneous tissue disorders'

(e.g., Stevens Johnson syndrome, acute generalized exanthematous pustulosis), 'immune system disorders' (e.g., anaphylaxis, hypersensitivity reactions), 'nervous system disorders' (e.g., seizures, serotonin syndrome), and 'hepatobiliary disorders' (e.g., drug-induced liver injury, hepatitis reactivation). The highest-level labeling recommendations from the reviews included modifications to the Boxed Warning (2), Warnings and Precautions (33), Contraindications (2), Adverse Reactions (30), and Drug Interactions (2).

Characteristics of the 10,381 ICSRs included for modeling are provided in Table 2 by ICSR inclusion (PVU=1) or exclusion status (PVU=0). Overall, 12.5% of ICSRs did not report a serious outcome. Included and excluded ICSRs provided a similarly high proportion of complete demographic information (age ~86%, sex ~94%). Data fields with a 5%-or-more absolute percentage difference in complete information in the included group were weight (35.5% vs 29.1%), reason for use (82.7% vs 71.9%), dose (76.9% vs 69.4%), and start date (44.3% vs 37.0%). ICSRs with an outcome of death were less frequent in the included group (4.3% vs 12.4%). A higher proportion of included ICSRs reported a positive dechallenge (42.8% vs 25.8%), positive rechallenge (2.7% vs 1.5%), and DME (6.3% vs 3.7%). ICSRs with more than one suspect product were more frequent in the excluded group (48.6% vs 32.4%).

3.2 Predictive Model Performance and Validation

The parsimonious model was selected as best performing model for validation. The model containing only completeness variables had poorer discrimination than the full or parsimonious model (AUC 0.63 vs 0.707 and 0.706, respectively; Fig. 1) and poor calibration (Hosmer and Lemeshow $p<0.05$). While the full and parsimonious models had similar discrimination and calibration, the parsimonious model had a lower AIC. The strongest predictors of ICSR inclusion were reporting of a DME (OR 1.93, 95% CI 1.54–2.43), positive dechallenge (OR 1.67, 95% CI 1.50–1.87), and reason for product use provided (OR 1.60, 95% CI 1.40–1.83) (Table 3). The strongest predictors of ICSR exclusion were death reported as the only outcome (OR 2.70, 95% CI 1.69–4.35), more than three suspect products (OR 2.70, 95% CI 2.22–3.23), > 15 preferred terms (OR 2.70, 95% CI 1.79–3.85), and little narrative information defined as < 100 words in the narrative and no ICSR attachment (OR 2.16, 95% CI 1.56–3.07). The availability of information (i.e., populated data fields) was positively associated with ICSR inclusion. Additionally, ICSRs originating from the USA, non-consumer reporter, or classified as literature were also predictive of ICSR inclusion in a case series. In contrast, ICSRs with any serious outcome reported and expedited ICSRs were more likely to be excluded.

Table 1 Characteristics of the pharmacovigilance reviews in which individual case safety reports (ICSRs) were selected for modeling

	<i>N</i> =69 reviews
Number of ICSRs reviewed, mean (median); range	154.6 (89.5); 4–1376
Number of duplicate ICSRs identified, mean (median); range	17.3 (8); 0–190
Number of ICSRs excluded from case series, mean (median); range	125.6 (63); 1–1057
Number of ICSRs included in case series, mean (median); range	30.7 (16); 1–319
Product type, <i>n</i> (%)	
Biologic	8 (11.6)
Drug	61 (88.4)
Reviewed product therapeutic area (<i>n</i> > 1) ^a , <i>n</i> (%)	
Analgesia/musculoskeletal	3 (4.3)
Antiviral	4 (5.8)
Cardiology	11 (15.9)
Dermatology	4 (5.8)
Endocrine/metabolic	9 (13.0)
Gastrointestinal	8 (11.6)
Neurology	8 (11.6)
Oncology/hematology	11 (15.9)
Psychiatric	5 (7.2)
Reproductive	3 (4.3)
Products reviewed, <i>n</i> (%)	
Single product	55 (79.7)
Class(es)	12 (17.4)
Drug–drug interaction	2 (2.9)
Primary reviewed event system organ class (<i>n</i> > 1) ^b , <i>n</i> (%)	
Blood and lymphatic system disorders	2 (2.9)
Cardiac disorders	5 (7.2)
Gastrointestinal disorders	5 (7.2)
General disorders and administration site conditions	4 (5.8)
Hepatobiliary disorders	6 (8.7)
Immune system disorders	8 (11.6)
Infections and infestations	2 (2.9)
Investigations	2 (2.9)
Metabolism and nutrition disorders	2 (2.9)
Musculoskeletal and connective tissue disorders	4 (5.8)
Nervous system disorders	7 (10.1)
Psychiatric disorders	2 (2.9)
Renal and urinary disorders	2 (2.9)
Respiratory, thoracic and mediastinal disorders	2 (2.9)
Skin and subcutaneous tissue disorders	12 (17.4)
Vascular disorders	3 (4.3)
Highest labeling recommendation, <i>n</i> (%)	
Boxed warning	2 (2.9)
Warnings and precautions	33 (47.8)
Contraindication	2 (2.9)
Adverse reactions	30 (43.5)
Drug interactions	2 (2.9)

^aAllergy, bone, medical imaging, nutrition, rheumatologic, urinary products were evaluated in one review each

^bEye disorder, neoplasms benign, malignant and unspecified, product issue and reproductive system and breast disorders were evaluated in one review

Table 2 Characteristics of individual case safety reports (ICSRs) excluded (PVU = 0) and included (PVU = 1) from the review pool case series

Characteristics ^a	PVU=0 (N=8266)	PVU=1 (N=2115)	All ICSRs (N=10,381)
Report type (%)			
Direct	5.6	7.6	6.0
Expedited (15-day)	74.8	65.1	72.8
Non-expedited	19.7	27.3	21.2
Reporter (%)			
Consumer	25.8	19.8	24.6
Healthcare professional	71.8	78.7	73.2
Unknown or other	2.3	1.5	2.2
Age provided (%)	86.1	85.7	86.0
Sex provided (%)	93.7	95.7	94.1
Missing both age and sex (%)	5.1	3.4	4.8
Weight provided (%)	29.1	35.5	30.4
Reason for use provided (%)	71.9	82.7	74.1
Dosage/dose provided (%)	69.4	76.9	70.9
Product manufacturer provided (%)	54.6	58.7	55.4
Product lot provided (%)	9.1	10.5	9.4
Product NDC provided (%)	0.8	0.9	0.8
Product start date provided (%)	37.0	44.3	38.5
Product stop date provided (%)	26.4	36.1	28.4
Event date provided (%)	65.1	67.9	65.6
Any serious outcome (%)	88.7	82.7	87.5
Death	12.4	4.3	10.7
Hospitalization	44.5	38.4	43.3
Life threatening	8.2	7.0	7.9
Disability	3.7	4.8	3.9
Congenital anomaly	0.2	0.0	0.1
Other serious	49.9	49.8	49.9
Death only outcome (%)	4.3	1.0	3.6
Other serious only outcome (%)	33.3	36.0	33.8
> 1 Serious outcome (%)	24.8	19.2	23.7
Curated preferred terms (%)	23.0	17.9	21.9
Curated narrative words (%)	26.2	27.2	26.4
Follow-up report provided (%)	36.4	38.4	36.8
Designated medical event ^b (%)	3.7	6.3	4.2
Positive dechallenge (%)	25.8	42.8	29.2
Positive rechallenge (%)	1.5	2.7	1.7
Limited narrative (%)	3.9	2.1	3.5
Literature (%)	15.4	15.0	15.4
USA (%)	61.6	66.6	62.6
Number of suspects (%)			
1	51.4	67.6	54.7
2–3	28.4	24.9	27.7
≥4	20.1	7.5	17.6
Number of preferred terms (%)			
1–5	68.7	80.0	71.0
6–15	25.0	18.2	23.6
> 15	6.3	1.8	5.4

NDC national drug code, PVU pharmacovigilance utility

^aSee Supplementary Materials Table 1 in the electronic supplementary material for operational definitions

^bSee Supplementary Materials Table 2 in the electronic supplementary material for designated medical event list

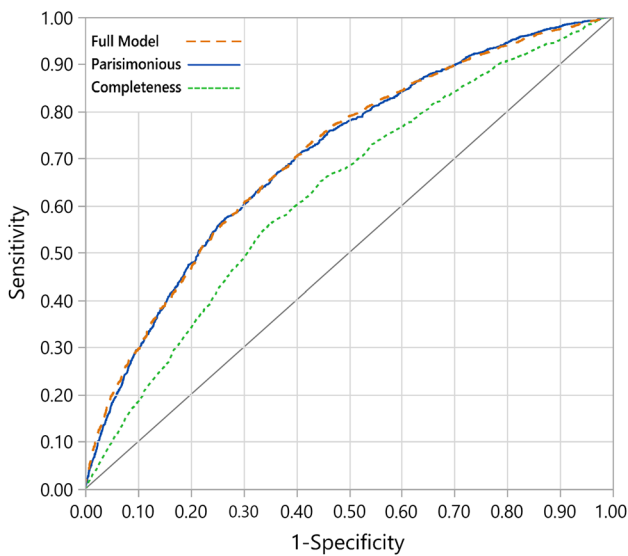


Fig. 1 Receiver operating curves for the full, parsimonious, and completeness models

The Hosmer and Lemeshow goodness-of-fit test indicated acceptable calibration ($p=0.22$). The optimism-adjusted C -statistic was 0.705, indicating minimal overfitting of the model. The GEE model provided the same estimated adjusted odds ratios for the parameters with wider confidence intervals. The GEE model also had the same discriminatory ability with an AUC of 0.702 (Supplementary Materials Table 3, see ESM).

3.3 Model Performance Across Safety Issues

We performed a sensitivity analysis to evaluate the performance of the validated model across different safety issues. ICSRs from reviews evaluating hypersensitivity/anaphylaxis, drug-induced liver injury, cardiovascular events (including ventricular arrhythmias, torsades de pointes, QT prolongation, hypotension, AV block, heart failure), and events without acute life-threatening outcomes (including alopecia, application-site pigmentation changes/scarring, false-positive drug screen, fingerprint loss, weight gain) were selected for this analysis. Discriminative performance of the model within these review issues is summarized in Table 4 and illustrated in Fig. 2. The correct classification of cases was higher when the model evaluated hypersensitivity reactions and drug-induced liver injury ($c=0.70$ and 0.74 , respectively); however, the model demonstrated lower discrimination with cardiovascular events and events without acute life-threatening outcomes ($c=0.64$ and 0.58 , respectively).

Table 3 Association between factors and inclusion of individual case safety reports (ICSRs) in a case series (parsimonious model^a, $c=0.71$)

Variables	Odds ratio	95% Confidence interval
Missing both age and sex	0.79	0.61–1.04
Weight provided	1.32	1.17–1.48
Reason for use provided	1.60	1.40–1.83
Dose/dosage provided	1.12	0.99–1.27
Product manufacturer provided	1.12	1.01–1.24
Product NDC provided	0.62	0.35–1.09
Product stop date provided	1.45	1.29–1.64
Serious outcome	0.51	0.42–0.62
Death outcome only	0.37	0.23–0.59
Other serious outcome only	1.11	0.99–1.25
Designated medical event ^b reported	1.93	1.54–2.43
Positive dechallenge	1.67	1.50–1.87
Positive rechallenge	1.33	0.95–1.87
Limited narrative	0.46	0.33–0.64
Report type (vs non-expedited)		
Direct report	1.11	0.87–1.41
Expedited reports	0.84	0.72–0.98
USA report	1.24	1.10–1.40
Consumer report	0.60	0.53–0.69
Literature report	1.34	1.15–1.57
Suspects reported (vs 1)		
2–3	0.70	0.63–0.79
> 3	0.37	0.31–0.45
Preferred terms reported (vs 1–5)		
6–15	0.74	0.65–0.84
> 15	0.37	0.26–0.52

NDC national drug code

^aParsimonious model developed via fast backward technique that retained all variables with a p value <0.15

^bSee Supplementary Materials Table 2 in the electronic supplementary material for designated medical event list

4 Discussion

We developed and validated a model predictive of ICSRs selected by FDA analysts from FAERS that ultimately informed a regulatory action. To our best knowledge, no studies have attempted to develop a model predictive of an ICSR’s PVU using a surrogate of ICSR inclusion in a case series supporting a regulatory recommendation. This practical outcome enabled the collection of a large and diverse ICSR training set for model development. As a result, our model considered the incorporation of many potential predictors, including those that are relatively uncommon in FAERS such as presence of a positive rechallenge. Our analysis suggests that the use of completeness elements alone to

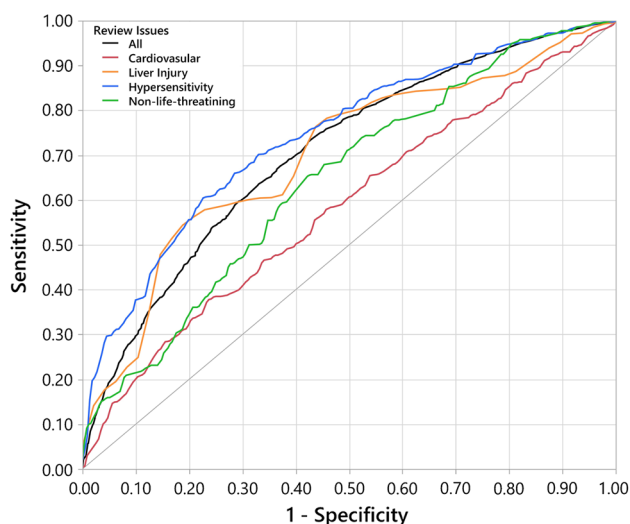
Table 4 Model performance by different review issue subsets

Review subsets	Reviews (<i>N</i>)	ICSRs (<i>N</i>)	Included ICSRs (<i>N</i>)	AUC
All reviews	69	10,381	2115	0.706
Cardiovascular events ^a	6	949	232	0.582
Drug-induced liver injury	6	1135	48	0.701
Hypersensitivity/anaphylaxis	7	1374	221	0.742
Non-acutely-life-threatening events ^b	8	658	256	0.644

AUC area under the curve, ICSRs individual case safety reports

^aSafety issues included torsades de pointes, QT prolongation, severe ventricular arrhythmias, hypotension, syncope, heart failure

^bSafety issues included alopecia, application-site pigmentation changes/scarring, false-positive drug screen, fingerprint loss, weight gain

**Fig. 2** Receiver operating curves by safety issue

predict PVU had limited discriminative ability but was better than random. Supplementing completeness elements with other aspects of ICSRs such as the number of suspect products resulted in significant improvements in discrimination.

The determinants of PVU evaluated generally performed as expected. The availability of patient and product information was positively associated with ICSR inclusion. Additionally, presence of a positive dechallenge, positive rechallenge, and DME were positively associated with inclusion in a case series. The presence of a DME is not surprising, as these are specific severe events of interest (e.g., toxic epidermal necrolysis) rather than imprecise and often non-serious events (e.g., rash or erythema) that may be retrieved in a reviewer's broader search. Importantly, ICSRs without a DME can still have a higher predicted likelihood of case inclusion than a DME containing ICSR if other important predictors are present in the non-DME ICSR but absent in the DME containing ICSR. We also conducted an additional analysis that excluded the DME variable and found removal from the model did not change the *c*-statistic significantly

(i.e., 0.71 with the DME variable included to 0.70 without the DME variable). ICSRs from the US and from the medical literature were also positively associated with ICSR inclusion. Publication of individual case reports or case series in the medical literature clearly has a higher bar for detail than entry of an ICSR into FAERS, for which only minimal criteria exist [15]. Our finding also indicated that reporting more than three suspect products was a strong predictor of ICSR exclusion, which likely reflects the difficulty in separating the effect of one drug from that of the others in assessing causality. ICSRs reporting death as the only outcome was also a strong predictor of ICSR exclusion. While this finding may seem counterintuitive, we examined the data further and found that one-third of death-only ICSRs were associated with the American Association of Poison Control Centers. These ICSRs tend to describe multiple-substance overdoses and their clinical consequences [16]. Overdose was not the subject of any of the reviews included in the model development pool, but these cases were retrieved because they were coded with other AEs (e.g., liver injury, arrhythmias).

Our evaluation of the model's ability to correctly classify ICSRs within different subgroups of events highlights opportunities to develop algorithms tailored to safety issues. The most frequent safety issues evaluated in reviews were hypersensitivity/anaphylactic reactions and drug-induced liver injury. Thus, it would not be unexpected for the model to perform better for these subgroups of focused safety issues relative to subgroups of reviews that evaluated broader topics. Logically, important predictors of an ICSR's utility will vary by safety issue. For example, the presence of electrocardiogram information or electrolyte laboratory results contained in an ICSR may have significant predictive ability for identifying the most pertinent ICSRs for torsades de pointes, and little to no predictive value for events like serious skin reactions. Natural language processing methods could be leveraged to identify these potentially predictive variables currently captured as free-text in the narrative [17]. While we had sufficient sample size to develop a stable model across the considered case series, we could not properly

address heterogeneity regarding predictors. Larger studies capturing a broader and more representative range of case series allowing targeted analyses would be needed. Given a large proportion of reviews were excluded, additional work would also be needed to further evaluate the generalizability. Similarly, pharmacovigilance experts use sources of safety knowledge external to the FAERS database in their assessments of ICSRs. Further development of models that leverage additional data sources (e.g., product labeling, case definitions, pharmacology information, molecular pathways) should enhance a model's predictive abilities [18].

Because our model was developed on ICSRs in which a signal was already identified, the most direct application is to prioritize ICSRs for manual review once a signal is identified. However, a logical extension may be the prioritization of ICSRs for screening independent of a signal. Implementation would require consideration of each variable's applicability in the context of signal identification practices and subsequent model validation. For example, if ICSRs coded with DME preferred terms were already automatically triaged for screening, there is no value in including this variable. Further, utility predictions on the ICSR level could be used to enhance approaches currently used to identify signals at the product-event level. For example, Caster et al. developed *vigiRank*, a method that combines multiple strength-of-evidence aspects including disproportionality and completeness of individual ICSRs within the product–event pairs evaluated in *vigiBase* [19]. The Netherlands Pharmacovigilance Centre Lareb recently developed a similar method with different predictors [20]. In both cases, combining disproportionality measures at the product–event level with aspects of the underlying ICSRs resulted in increased performance in signal detection compared with disproportionality analyses alone [20, 21]. Future work should consider the development and validation of these approaches for signal detection in FAERS.

Meaningful measures of utility are challenging to operationalize and represent an area of much needed research [22, 23]. While our PVU outcome represents a contribution, there are several notable limitations. Dichotomizing each ICSR as only valuable if it supports a regulatory action oversimplifies the contributions of ICSRs to safety knowledge. Within a case series certain ICSRs could be considered more valuable, such as those determined to have a high likelihood of causality or those illustrating unique aspects. It is possible that a case classified as $PVU = 0$ in the context of a particular review may be classified as $PVU = 1$ for a review of a different product or event. This possibility exists because multiple products and events may be reported within a single ICSR. For example, an ICSR may describe both a rash with a suspect product and liver failure. While this ICSR may be classified as $PVU = 0$ for a review evaluating serious skin reactions, it may have $PVU = 1$ in a future review focused on

the evaluation of drug-induced liver injury. Additionally, the search strategies utilized to obtain the ICSRs for review may vary depending on the particular circumstance. It is possible that more specific searches may have a higher yield than those with broader criteria (e.g., a search for cases coded with anaphylactic reaction only versus a search including any terms associated with anaphylactic reactions like hypotension), but we were unable to account for the underlying rationale for each search's criteria. Finally, the inclusion of an ICSR in a case series is a structured but subjective process based on the reviewer's expertise, case definition utilized, and assessment of causality [24]. Agreement on causality using various available causality tools has demonstrated variability [25].

Finally, duplicative reporting presents a universal challenge for most public health surveillance systems. While we used the latest version of an ICSR for model development, ICSRs describing the same occurrence of an adverse event in a patient reported by more than one reporter can result in duplicates. The reporters could be manufacturers of the different products when there are multiple suspect products, different manufacturers of a signal suspect product (e.g., brand and generic manufacturers), or reports submitted directly by the public to FAERS that were also submitted by a manufacturer. As a result, a particular patient–drug–event may have been reported by three separate manufacturers, resulting in three unique ICSRs with the same content. The FDA analyst may have included one ICSR in the case series; excluding the two duplicates, which would result in contradictory classifications for the same ICSR. The reason for exclusion was not delineated for each ICSR in most reviews; therefore, we were unable to account for decisions made with duplicates involved. This misclassification can impact the predictive ability of the model. This study has highlighted the need for better accounting of excluded duplicates as well as quantifying their impact. While duplicate ICSRs are often cited as a limitation of FAERS, the extent may be larger than some realize. We examined the data further and found that over 50% (39/69) of reviews identified at least 10% of ICSRs as duplicates and 23% (16/69) identified more than 25%. Reliable deduplication is still a challenging aspect of pharmacovigilance [26].

5 Conclusion

Our study demonstrated the feasibility of developing predictive tools to augment review of ICSRs from FAERS. The model's modest discriminative ability highlights opportunities for further enhancement and suggests algorithms tailored to safety issues may be beneficial.

Compliance with Ethical Standards

Funding No funding was used for the preparation of this manuscript.

Conflict of interest Monica A. Muñoz, Gerald J. Dal Pan, Jenny Wei, Chris Delcher, Hong Xiao, Cindy M. Kortepeter, and Almut G. Winterstein declare that they have no conflict of interest.

Ethical approval This study was approved by both the US FDA and University of Florida's Institutional Review and Boards.

References

- Lester J, Neyarapally GA, Lipowski E, Graham CF, Hall M, Dal Pan G. Evaluation of FDA safety-related drug label changes in 2010. *Pharmacoepidemiol Drug Saf.* 2013;22(3):302–5.
- Ishiguro C, Hall M, Neyarapally GA, Dal Pan G. Post-market drug safety evidence sources: an analysis of FDA drug safety communications. *Pharmacoepidemiol Drug Saf.* 2012;21(10):1134–6.
- FDA Adverse Event Reporting System (FAERS) Public Dashboard. <https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070093.htm>. Accessed 6 Dec 2019.
- Duggirala, HJ, Tønning JM, Smith E, Bright RA, Baker JD, Ball R, et al. Data mining at FDA. <http://www.fda.gov/downloads/ScienceResearch/DataMiningatFDA/UCM443675.pdf>. Accessed 28 Nov 2019.
- Bergvall T, Noren GN, Lindquist M. *vigiGrade*: a tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Saf.* 2014;37(1):65–77.
- Guidance for Industry. Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment, US Food and Drug Administration; 2005. <http://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126834.pdf>. Accessed 28 Nov 2019.
- FDA Drug Safety Guidances. <https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs>. Accessed 28 Nov 2019.
- Moore TJ, Furberg CD, Mattison DR, Cohen MR. Completeness of serious adverse drug event reports received by the US Food and Drug Administration in 2014. *Pharmacoepidemiol Drug Saf.* 2016;25(6):713–8.
- Oosterhuis I, Rolfes L, Ekhardt C, Muller-Hansma A, Harmark L. First experiences with a tool to measure the level of clinical information present in adverse drug reaction reports. *Expert Opin Drug Saf.* 2018;17(2):111–5.
- Han L, Ball R, Pamer CA, Altman RB, Proestel S. Development of an automated assessment tool for MedWatch reports in the FDA adverse event reporting system. *J Am Med Inform Assoc.* 2017;24(5):913–20.
- CIOMS Working Group VIII. Practical aspects of signal detection in pharmacovigilance. Geneva: Report of CIOMS Working Group VIII; 2010.
- Pinnow E, Amr S, Bentzen SM, Brajovic S, Hungerford L, St George DM, et al. Postmarket safety outcomes for new molecular entity (NME) drugs approved by the food and drug administration between 2002 and 2014. *Clin Pharmacol Ther.* 2018;104(2):390–400.
- Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc.* 1983;78:316–31.
- Steyerberg EW, Harrell FE Jr, 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(8):774–81.
- Postmarketing reporting of adverse drug experiences. 21 CFR 314.80.
- Gummin DD, Mowry JB, Spyker DA, Brooks DE, Fraser MO, Banner W. 2016 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 34th Annual Report. *Clin Toxicol (Phila).* 2017;55(10):1072–252.
- Wunnava S, Qin X, Kakar T, Socrates V, Wallace A, Rundensteiner E. Towards transforming FDA adverse event narratives into actionable structured data for improved pharmacovigilance. In: 2017 Proceedings of the symposium on applied computing, pp 777–782. <https://doi.org/10.1145/3019612.3022875>.
- Schotland P, Racz R, Jackson D, Levin R, Strauss DG, Burkhardt K. Target-adverse event profiles to augment pharmacovigilance: a pilot study with six new molecular entities. *CPT Pharmacomet Syst Pharmacol.* 2018;7(12):809–17.
- Caster O, Juhlin K, Watson S, Noren GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in *vigiRank*. *Drug Saf.* 2014;37(8):617–28.
- Scholl JHG, van Hunsel F, Hak E, van Puijenbroek EP. A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol Drug Saf.* 2018;27(2):199–205.
- Caster O, Sandberg L, Bergvall T, Watson S, Noren GN. *vigiRank* for statistical signal detection in pharmacovigilance: first results from prospective real-world use. *Pharmacoepidemiol Drug Saf.* 2017;26(8):1006–10.
- Dal Pan GJ. Ongoing challenges in pharmacovigilance. *Drug Saf.* 2014;37(1):1–8.
- PRAC Strategy on Measuring the Impact of Pharmacovigilance Activities. European Medicines Agency. November 30, 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/01/WC500199756.pdf. Accessed 18 Apr 2018.
- Meyboom RH, Hekster YA, Egberts AC, Gribnau FW, Edwards IR. Causal or casual? The role of causality assessment in pharmacovigilance. *Drug Saf.* 1997;17(6):374–89.
- Agbabiaka TB, Savovic J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. *Drug Saf.* 2008;31(1):21–37.
- Noren GN. The power of the case narrative—can it be brought to bear on duplicate detection? *Drug Saf.* 2017;40(7):543–6.