



Using an Automated Algorithm to Identify Potential Drug-Induced Liver Injury Cases in a Pharmacovigilance Database

Liliam Pineda Salgado · Ritu Gupta · Michael Jan ·
Osman Turkoglu · Alvin Estilo · Vinu George · Mirza I. Rahman

Received: May 31, 2021 / Accepted: July 6, 2021 / Published online: July 28, 2021
© The Author(s) 2021

ABSTRACT

Introduction: Drug-induced liver injury (DILI) is the most frequent cause of acute liver failure in North America and Europe, but it is often missed because of unstandardized diagnostic methods and criteria. This study aimed to develop and validate an automated algorithm to identify potential DILI cases in routine pharmacovigilance (PV) activities.

Methods: Post-marketing hepatic adverse events reported for a potentially hepatotoxic drug in a global PV database from 19 March 2017 to 18 June 2018 were assessed manually and with the automated algorithm. The algorithm provided case assessments by applying pre-specified criteria to all case data and narratives simultaneously.

Results: A total of 1456 cases were included for analysis and assessed manually. Sufficient data for algorithm assessment were available for 476

cases (32.7%). Of these cases, manual assessment identified 312 (65.5%) potential DILI cases while algorithm assessment identified 305 (64.1%) potential DILI cases. Comparison of manual and algorithm assessments demonstrated a sensitivity of 97.8% and a specificity of 79.3% for the algorithm. Given the prevalence of potential DILI cases in the population studied, the algorithm was calculated to have positive predictive value 56.3% and negative predictive value 99.2%. The time required for manual review compared to algorithm review suggested that application of the algorithm prior to manual screening would have resulted in a time savings of 42.2%.

Conclusion: An automated algorithm to identify potential DILI cases was developed and successfully implemented. The algorithm demonstrated a high sensitivity, a high negative predictive value, along with significant efficiency and utility in a real-time PV database.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12325-021-01856-x>.

Keywords: Automated algorithm; Drug-induced liver injury; Drug safety; Hepatotoxicity; Pharmacovigilance

L. Pineda Salgado (✉) · R. Gupta · M. Jan ·
O. Turkoglu · A. Estilo · V. George · M. I. Rahman
Otsuka Pharmaceutical Development &
Commercialization, Inc., 508 Carnegie Center,
Princeton, NJ 08540, USA
e-mail: Liliam.PinedaSalgado@otsuka-us.com

R. Gupta
Vitrana Inc., Cranbury, NJ 08512, USA

Key Summary Points

An automated algorithm was developed and validated for identification of potential DILI cases in a real-time, real-world PV database.

The algorithm was designed and optimized to maximize inclusion of potential DILI cases.

The algorithm demonstrated a sensitivity of 97.8% and a specificity of 79.3%.

Compared to manual case review, application of the automated algorithm resulted in an estimated time saving of 42.2%.

INTRODUCTION

Drug-induced liver injury (DILI) is a potentially fatal adverse drug reaction that is the most frequent cause of acute liver failure in North America and Europe [1–3]. Concern for DILI is one of the main barriers to marketing authorization, one of the most frequent causes of post-marketing restrictions, and one of the major reasons for marketing withdrawal [4–8]. Despite significant downstream morbidity and mortality, DILI eludes early upstream detection because associated signs and symptoms are nonspecific and pathognomonic biomarkers remain unelucidated [3, 9–11]. As there is no standardized method to predict DILI, creation of a systematic approach for early detection is currently the best strategy for prevention and potential intervention [1, 12].

While several criteria have been proposed to detect potential DILI cases, there is significant variability in how these criteria can be applied in stopping rules for clinical trial settings, risk–benefit assessments in post-marketing settings, reporting to health authorities, and to physicians and patients [1, 12–14]. Further, since most criteria involve a combination of

laboratory values and patient symptoms, manual review may be needed in addition to automated filtering by laboratory criteria [15, 16]. A recent meta-analysis of algorithms created to identify potential DILI cases found a low range of detection (1.0–40.2%) that varied with threshold criteria, case definitions, diagnostic codes, and study drugs [17]. Many of these algorithms were applied to retrospective data, and there remains a significant need to develop more efficient automated methods to consider cases prospectively as part of active, ongoing pharmacovigilance surveillance.

Given the lack of standardized methods to routinely monitor for DILI, we developed an automated algorithm to facilitate detection of potential DILI cases. The algorithm does not confirm diagnosis of DILI. The objective of this study was to evaluate the application of our potential DILI detection algorithm in activities related to routine pharmacovigilance. The logic of the algorithm can be further applied to identify other criteria-based pathologies.

METHODS

Datasets and Case Selection

The datasets used in this study compiled post-marketing individual case safety reports (ICSRs) from a real-time, global PV database between 19 March 2017 and 18 June 2018. Thirteen of 15 datasets corresponding to monthlong reporting periods were included in the final analysis as algorithm assessment was not performed for two non-consecutive monthlong reporting periods.

To identify potential DILI cases among reported hepatic adverse events, dataset inclusion criteria consisted of (1) initial or follow-up ICSR reporting for a specific potential hepatotoxic agent during the respective monthlong period and (2) at least one lower level MedDRA (Medical Dictionary for Regulatory Activities) term contained within the following five hepatic Standardized MedDRA Queries (SMQs) [18]:

1. Cholestasis and jaundice of hepatic origin

2. Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions
3. Hepatitis, non-infectious
4. Liver-related investigations, signs, and symptoms
5. Liver-related coagulation and bleeding disturbances

As the global PV database is maintained and updated in real time, the same ICSR could appear in successive datasets as additional follow-up information was received and case details and assessment evolved. While the study period spanned MedDRA versions 20.0, 20.1, and 21.0, SMQ parameters were maintained throughout versioning by the MedDRA Maintenance and Support Services Organization [18].

Case Assessment

In practice, pharmacovigilance activities related to potential hepatotoxic agents include identification of potential DILI cases that will then be further evaluated by a hepatic adjudication committee, consisting of an expert panel of hepatologists. The case reviews and analyses performed in this study occurred prior to submission to a hepatic adjudication committee, reflecting the process of deciding which cases describe potential DILI and merit hepatic adjudication. For the purposes of this study cases identified as potential DILI cases through manual review were considered to be “positive” and were sent for additional assessment by a hepatic adjudication committee, whereas cases that were not identified as potential DILI cases were considered to be “negative” and not sent for hepatic adjudication. Assessments performed in this study do not reflect the judgment of a hepatic adjudication committee, nor do they confirm diagnosis of DILI.

Conservatively expanding laboratory ranges from the US Food and Drug Administration (FDA) guidance for DILI evaluation to be inclusive, increases of serum alanine transaminase (ALT), serum aspartate transaminase (AST), serum total bilirubin (TB), or international normalized ratio of prothrombin time (INR) falling into any of the following ranges were

considered sufficient to identify a potential DILI case [12]:

- ALT or AST $\geq 7.5 \times \text{ULN}^*$
- ALT or AST $\geq 5.0 \times \text{ULN}$ for more than 2 weeks
- ALT or AST $\geq 3.0 \times \text{ULN}$ and (TB $\geq 2.0 \times \text{ULN}$ or INR ≥ 1.5)
- ALT or AST $\geq 3.0 \times \text{ULN}$ with the appearance of fatigue, nausea, vomiting, abdominal pain upper, fever, rash, and/or eosinophilia ($> 5.0\%$)

* $\times \text{ULN}$ denotes the proportion of how many times the laboratory value is above the upper limit of its normal reference range [19]

The FDA guidance also suggests surveillance of cases that meet the following [12]:

- ALT or AST $\geq 2.0 \times \text{ULN}$ or twofold increases above baseline values for subjects with elevated values before drug exposure

Cases meeting any of the aforementioned criteria were assessed as “positive” for potential DILI.

Manual Case Review

Once each dataset was generated, all cases in the dataset were manually reviewed and analyzed over a 2-week period. The manual reviewer was the same individual for all datasets to maintain consistency in case review, analysis, and assessment throughout the study.

Assessment of each case consisted of applying medical judgment to evaluate laboratory data as well as medical information in the case narrative. ALT, AST, TB, and INR values were evaluated in terms of the proportion of how many times they were elevated above the upper limit of the respective normal value ranges provided. When normal value ranges were not provided, 40 international units per liter (IU/L) was used as the ULN for ALT and AST, 1 mg per deciliter (mg/dL) was used as the ULN for TB, and 1 was used as the ULN for INR. Elevations from ULN as well as elevations from baseline were calculated. Cases that met the selection criteria described in the “[Case Assessment](#)” section were assessed as “positive” for potential

DILI. Cases that did not meet the selection criteria were assessed as “negative” for potential DILI.

Algorithm Development and Case Review

The categories of laboratory value ranges used above for manual case assessment were used to design and develop an algorithm to identify potential DILI cases. The algorithm was programmed using procedural language for structured query language (PL/SQL) to analyze PV data stored in Oracle System Tables (Oracle Corporation, Redwood City, CA). From user-specified inputs of time period, drug, SMQs, and ICSR source (e.g., clinical trial, post-marketing), the algorithm identified a subset of the global PV database upon which to apply selection criteria. This initial process recreated the same datasets that were assessed manually; however, as the algorithm was applied to a real-time PV database, data were current up to the time of the algorithm run, which occurred after manual review.

Next, the algorithm applied the case selection criteria itemized in the “[Case Assessment](#)” section. Using case data entered into laboratory, event, narrative, and date fields, the algorithm simultaneously identified which criteria, if any, were satisfied by the cases. ALT, AST, TB, and INR values were compared to respective ULNs and narratives were scanned for keywords fatigue, nausea, vomiting, abdominal pain upper, fever, rash, and/or eosinophilia.

For assessment, the algorithm designated a case as “positive” for potential DILI if it determined that the case met any one of the selection criteria categories previously described. The algorithm assessed cases as “negative” for potential DILI if the selection criteria were not met. Throughout the course of the study, several practical optimizations and modifications were made to the algorithm in order to improve its accuracy and ability to detect specified selection criteria. These optimizations included improvements of the user-facing output to delineate exact criteria met for potential DILI identification and exact \times ULN of laboratory value elevation.

Cases lacking laboratory values for ALT, AST, TB, and INR were unable to be assessed by the algorithm as these laboratory values are part of the selection criteria that the algorithm was programmed to identify. Such cases were assessed only by manual review.

Efficiency Evaluation

To evaluate the standard practice of manual case review, a sample of “positive” and “negative” potential DILI cases were assessed by the manual reviewer and the time required for review and analysis was determined for each assessment. This was used to determine the average amount of time required to manually review “positive” and “negative” potential DILI cases, respectively.

To evaluate the effect on time savings of prescreening the dataset with the algorithm, a sample of “positive” and “negative” potential DILI cases as assessed by the algorithm was then assessed manually, with the manual reviewer being aware of the algorithm outcome beforehand. This was used to determine the average amount of time required to review prescreened “positive” and “negative” potential DILI cases, respectively.

The difference in time required for standard manual review as compared with manual review following prescreening with algorithm review was determined as a measure of time difference. Time difference as a proportion of time required for standard manual review was used to evaluate efficiency as in Eq. (1):

$$\text{Efficiency} = \frac{\text{Time}_{\text{Manualreview}} - \text{Time}_{\text{Prescreenedmanualreview}}}{\text{Time}_{\text{Manualreview}}} \quad (1)$$

Statistical Analyses

To calculate the sensitivity and specificity of this algorithm case assessment, the manual case assessment was used as the comparator gold standard since it is the only standard method to perform this surveillance pharmacovigilance activity. The positive likelihood ratio (LR+) and

negative likelihood ratio (LR⁻) were then derived from calculated sensitivity and specificity. Calculations for true positive rate (TPR) and false positive rate (FPR) were used to generate a receiver operating characteristic (ROC) curve. The trapezoidal rule was used to calculate the area under the ROC curve (AUROC) [20].

For the population sampled in this study, the prevalence of potential DILI cases was determined from manual case review as the proportion of cases assessed that were positive for potential DILI. Using this as the population prevalence of potential DILI, the positive predictive value (PPV) and negative predictive value (NPV) of the algorithm were calculated per standard practice [21]. Accuracy of algorithm performance was calculated in terms of overall percentage agreement with manual review [22]. All statistical analyses were performed using Microsoft Excel, v.16.0 (Microsoft Corporation, Redmond, WA).

Ethics Compliance

This study is based on previously reported data and does not contain any new studies with human participants or animals performed by any of the authors. Permission to access and analyze deidentified data was granted by Otsuka Pharmaceutical Development & Commercialization, Inc. (Princeton, NJ).

RESULTS

Patient Demographics

A total of 1456 cases were manually reviewed during 13 monthlong periods for detection of potential DILI in patients receiving the same potential hepatotoxic agent. These cases represent 719 unique patients with demographics described in Table 1. Demographic information is displayed by outcome of manual assessment. Of the total cases assessed, 312 of them (21.4%) were identified as potential DILI cases; 165 of these potential DILI cases (52.9%) occurred among female patients, while 462 of 1144 cases negative for potential DILI (40.4%) occurred

among female patients. Age distribution was similar for both positive and negative potential DILI cases with mean age 55 years and 52 years, respectively, and median age 52 years and 50 years, respectively. The majority of these cases from the study period were from Japan.

Algorithm Performance

The algorithm assessed 476 cases (32.7%) using selection criteria with laboratory values for ALT, AST, TB, and INR. Table 2 shows a comparison between manual case assessments and algorithm case assessments. On the basis of these comparisons, the algorithm was calculated to have a sensitivity of 97.8% and a specificity of 79.3%. Likelihood ratios were calculated as LR⁺ 4.73 and LR⁻ 0.03.

Of the six case assessments categorized as false negatives, five were due to a conservative lowering of thresholds during manual case assessment, and one was due to a concern for serious liver injury based on adverse event terms reported (Table S1 in the supplementary material). For the 43 case assessments categorized as false positives, 26 were due to serum enzyme elevations that did not persist for 2 weeks, nine were due to patient symptoms that did not correlate with enzyme elevations (lack of temporal association, confounding from non-hepatic adverse event), and eight were due to laboratory values being unavailable at the time of manual review that then subsequently became available at the time of algorithm review (Table S1 in the supplementary material).

Since analyses were performed in monthlong intervals, sensitivities and specificities were also calculated for each month of data reviewed (Table 3). These data were also used to generate an ROC curve with an AUROC of 0.95 (Fig. 1). Given that manual review identified 21.4% of cases as potential DILI, the positive predictive value of the algorithm for this population was calculated as 56.3% and the negative predictive value was calculated as 99.2%. Algorithm accuracy was calculated as 89.7% overall percentage agreement with manual review.

Table 1 Patient demographics of cases assessed manually

	Manual review	
	Positive	Negative
Patients (n)	312	1144
Gender		
Female	165	462
Male	147	615
Unknown	0	67
Age (n)	293	968
Mean (SD)	55 years (15 years)	52 years (14 years)
Median (IQR)	52 years (45–64 years)	50 years (42–60 years)
Range	18–95 years	8 months–100 years
COI		
Armenia	0	1
Austria	0	3
Belgium	1	16
Canada	6	203
China	0	4
France	3	22
Germany	10	55
Italy	3	6
Japan	275	744
Netherlands	0	1
Norway	5	0
South Korea	0	10
Spain	5	4
Switzerland	1	13
Thailand	0	3
UK	1	44
USA	2	15

COI country of incidence, *IQR* interquartile range, *SD* standard deviation

Table 2 Manual case assessments vs. algorithm case assessments

	Manual review	
	Positive	Negative
Algorithm review		
Positive	262 (TP)	43 (FP)
Negative	6 (FN)	165 (TN)

Sensitivity = 97.8%; Specificity = 79.3%; Positive predictive value = 56.3%; Negative predictive value = 99.2%; Overall percentage agreement = 89.7%

TP true positive, *FP* false positive, *FN* false negative, *TN* true negative

Using manual review outcomes as the gold standard for comparison, algorithm sensitivity and specificity were calculated. Positive predictive value and negative predictive value were calculated using an estimated population prevalence of 312/1456 potential DILI cases (21.4%) among ICSRs selected using 5 hepatic SMQs

Algorithm Efficiency

Manual review and analysis of cases that were identified as positive potential DILI cases required an average of 18.5 min per case for evaluation, whereas cases that were identified as negative, required an average of 7.5 min per case for evaluation. This difference was attributable to additional time being spent to calculate the extent of hepatic function test elevations, coupled with a narrative review to assess duration and associated symptoms.

The automated algorithm was able to provide assessments for all supplied cases within 1 s. Confirmation of the algorithm's assessments required an additional 1.5 min on average, for positive potential DILI cases only. Given the distribution of positive and negative cases in each monthlong assessment period and given that the algorithm was applied to cases with relevant laboratory values, Fig. 2 illustrates the potential time savings if the algorithm had been applied to screen all cases prior to manual review.

Overall, using the algorithm to screen all 1456 cases prior to manual review would be expected to save 101 h (42.2% of time

expended). Time savings per month ranged from 6.3 to 9.5 h with a time savings range of 35.4–47.7% in hours expended.

DISCUSSION

This study evaluated the application of an automated, multifactorial, algorithm in pre-screening ICSRs with hepatic adverse events, for identification of potential DILI cases in a real-time, real-world PV database. The algorithm was found to have a sensitivity of 97.8% and a specificity of 79.3% with an AUROC of 0.95. Moreover, given the high prevalence (21.4%) of potential DILI cases in the population studied, the algorithm demonstrated a positive predictive value of 56.3% and a negative predictive value of 99.2%. Application of the algorithm in prescreening datasets for potential DILI cases was estimated to save 42.2% of time expended from manual case review.

Despite the utility of the algorithm in facilitating identification of potential DILI cases, it must be restated that DILI itself remains difficult to predict. The liver functions at the intersection of numerous metabolic pathways and is equally subject to the effects of active metabolites as well as drugs, other agents, and the potential interactions between them [9, 23, 24]. Patient genetics, demographics, comorbidities, behavior, and environment all may play a role in precipitating DILI [9, 23, 25–27]. DILI may also occur with delayed onset, adding complexity to assessment of the temporal interplay between all these factors [28]. Thus, predicting DILI from a weighted analysis of risk factors remains difficult to validate consistently, though there has been much progress in the field. The focus of the current study was on the commonly reported pharmacovigilance information from which case assessments need to be made in deciding whether or not hepatic adjudication is required.

For pharmacovigilance purposes, it is essential that no potential DILI case is overlooked [1, 2, 9, 12]. To this end, the algorithm developed for this study was designed to optimize its sensitivity and negative predictive value. This is reflected in the algorithm's LR– of 0.03 and

Table 3 Case assessment by month

Month	Cases	Manual review	Algorithm review	True positives	False positives	False negatives	True negatives	Sensitivity (%)	Specificity (%)	Month
1	107	25	82	31	20	2	0	9	100.0	81.8
2	116	20	96	32	19	1	0	12	100.0	92.3
3	100	16	84	31	16	3	0	12	100.0	80.0
4	79	21	58	29	15	1	3	10	83.3	90.9
5	95	22	73	36	20	3	0	13	100.0	81.3
6	118	28	90	45	23	4	0	18	100.0	81.8
7	111	23	88	40	18	3	2	17	90.0	85.0
8	130	29	101	46	24	9	1	12	96.0	57.1
9	123	27	96	39	24	3	0	12	100.0	80.0
10	108	23	85	29	17	2	0	10	100.0	83.3
11	139	28	111	43	25	7	0	11	100.0	61.1
12	111	23	88	39	21	2	0	16	100.0	88.9
13	119	27	92	36	20	3	0	13	100.0	81.3

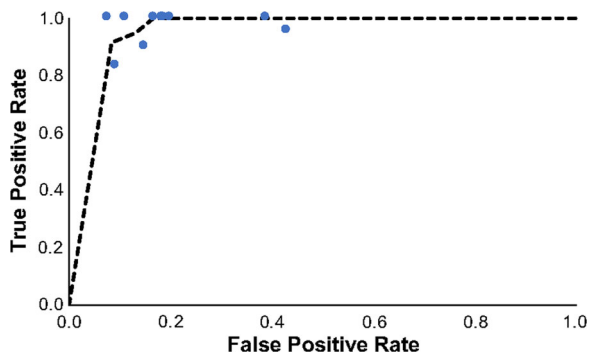


Fig. 1 Receiver operating characteristic curve for algorithm assessment. Receiver operating characteristic (ROC) curve (dashed line) constructed using monthly case assessments from Table 3 with monthly true positive rates and false positive rates plotted as ordered pairs (blue dots). Area under the ROC curve (AUROC) was calculated as 0.95 using the trapezoidal rule

LR+ of 4.73, where a “negative” potential DILI assessment would greatly decrease post-test probability and a “positive” potential DILI assessment would only moderately increase post-test probability by Bayes’ theorem [20].

However, it must be acknowledged that other algorithms and methods aimed at detecting potential DILI that optimize specificity do so by reducing sensitivity and increasing the risk of false negative, or missed potential DILI cases [17, 20–22, 29]. The algorithm was tested for 5 months beyond the study period to confirm accuracy and efficiency before it was incorporated into daily use. During this testing period and thereafter, algorithm assessments were made available during manual review, allowing for direct verification as opposed to independent comparison.

Time savings are a clear benefit of using automated methods and tools to perform pharmacovigilance activities. Though automation may never fully replace the ability to apply medical judgment in assessing case narratives and contextualizing laboratory findings, automation can usefully minimize the time spent performing repetitive comparisons and calculating laboratory values relative to set limits. Efforts to optimize this algorithm’s efficiency revealed areas for improvement in case

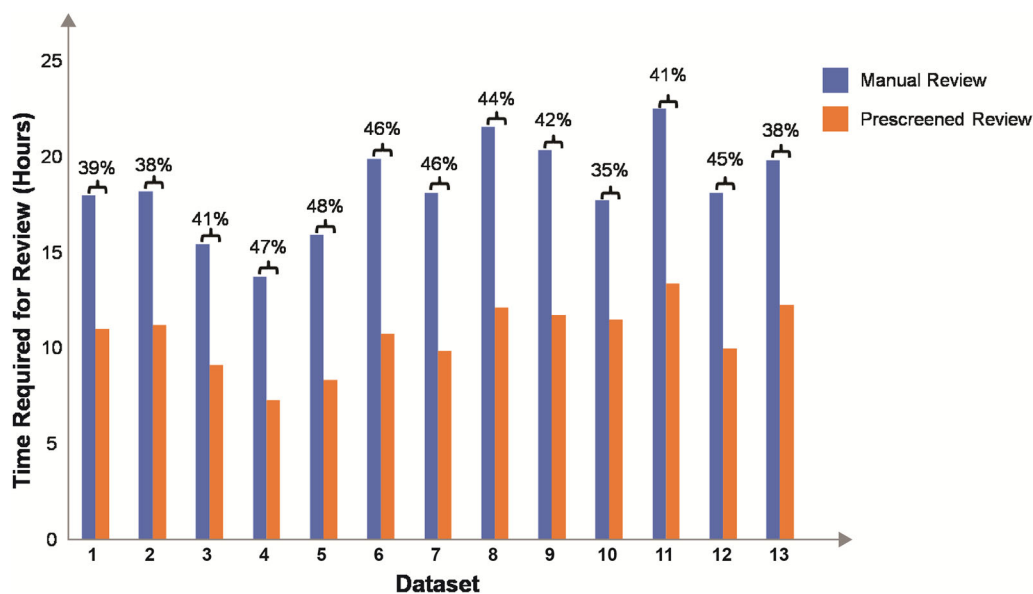


Fig. 2 Estimated time savings with algorithm review prior to manual review. Blue bars represent time required in hours for completion of manual case review and analysis. Orange bars represent estimated time required in hours for

completion of case review and analysis following prescreening with algorithm. Labeled values indicate time saved as percentage of total time expended for manual review and analysis

assessment methodology, as well as in case intake and case processing. The algorithm's dependence on proper entry of laboratory values identified opportunities to improve internal processes involved in ensuring that this happens.

Limitations of the algorithm included its inability to parse narrative information to correlate with laboratory findings. For example, a patient with ALT or AST elevation ≥ 3 ULN reporting "no fever" would have been identified by the algorithm as positive for potential DILI as the existence of the word "fever" in the narrative would have been interpreted as the patient having the symptom. One potential solution would be to incorporate natural language processing capabilities into the algorithm to interpret narrative context [30, 31]. Machine learning processes could provide additional input into making case assessments [32, 33].

Additionally, the algorithm was not always able to determine if the timing of symptoms corresponded to the timing of laboratory elevations. The algorithm also was not always able to determine duration of enzyme elevations. Operationally, one limitation of the preceding analysis was that the algorithm was often applied on a date after the initial period of manual review. The algorithm used case identifiers in the extracted dataset to assess corresponding laboratory values in the global safety database. As the global safety database is maintained in real time, on occasion the algorithm evaluated laboratory values that were not available at the time of manual review, resulting in a number of false positive case assessments. Finally, the datasets derived from ICSRs reporting hepatic adverse events would be expected to have a higher prevalence of potential DILI cases compared to the general population receiving a potential hepatotoxic agent [2, 3, 9]. As the algorithm was designed for analysis of a specific population with hepatotoxic injury and potential hepatotoxic agent use, its application to a generalized population would be difficult to interpret.

Nevertheless, the algorithm had many strengths in addition to its high sensitivity, high negative predictive value, and potential for significant time savings. The algorithm has no

dependency on software or MedDRA versioning, which makes it suitable to perform case assessment at any time past, present, or future [18]. The programming is also easily adaptable, allowing for changes in set thresholds as health authority guidance documents are updated.

As an example, the FDA's guidance for DILI evaluation in clinical trial cases is different from the post-marketing setting of this study, and the corresponding criteria could be easily programmed into the algorithm [12]. Though the algorithm's language processing was noted as a limitation, it is nevertheless a useful functionality that now has been adapted to identify serious liver injury event terms (e.g., hepatic failure, hepatitis fulminant, liver transplant) since the time of the study, further ensuring that potential DILI cases are not missed.

Finally, the algorithm may be adapted to identify other pathologies with multifactorial laboratory value selection criteria, such as drug reaction with eosinophilia and systemic symptoms (DRESS) syndrome, tumor lysis syndrome, neuroleptic malignant syndrome, and drug-induced renal injury [34–37].

CONCLUSION

We successfully developed and implemented a screening algorithm to assist in identifying potential DILI cases in support of routine pharmacovigilance activities. The algorithm demonstrated a high sensitivity, a high negative predictive value, along with significant efficiency and adaptability in a real-time PV database, which will ultimately result in cost savings. Notably, the algorithm's key features for pharmacovigilance purposes is a focus on sensitivity and negative predictive value, at the expense of specificity and positive predictive value, in our effort to maximize patient safety.

ACKNOWLEDGEMENTS

Funding. Otsuka Pharmaceutical Development & Commercialization, Inc. funded the

study, journal's Rapid Service fee and Open Access fee.

Authorship. All named authors meet the International Committee of Medical Journal Editors criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Author Contributions. Liliam Pineda Salgado, Ritu Gupta, Osman Turkoglu, Alvin Estilo, Vinu George, and Mirza I. Rahman conceived the idea and design of the project. Ritu Gupta developed and programmed the algorithm. Liliam Pineda Salgado acquired the data and performed all manual assessments. Liliam Pineda Salgado and Ritu Gupta compared manual assessments with algorithm assessments. Liliam Pineda Salgado, Ritu Gupta, and Michael Jan analyzed and interpreted the data. Liliam Pineda Salgado, Ritu Gupta, Michael Jan, Osman Turkoglu, Alvin Estilo, Vinu George, and Mirza I. Rahman made substantial contributions to drafting and revising the manuscript.

Disclosures. Ritu Gupta is an employee of Vitrana Inc. and a contractor for Otsuka Pharmaceutical Development & Commercialization, Inc. Liliam Pineda Salgado, Michael Jan, Osman Turkoglu, Alvin Estilo, Vinu George, and Mirza I. Rahman are employees of Otsuka Pharmaceutical Development & Commercialization, Inc.

Compliance with Ethics Guidelines. This study is based on previously reported data and does not contain any new studies with human participants or animals performed by any of the authors. Permission to access and analyze de-identified data was granted by Otsuka Pharmaceutical Development & Commercialization, Inc.

Data Availability. The data generated used for this study is the property of Otsuka Pharmaceutical Development & Commercialization, Inc. and is not publicly available.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Council for International Organizations of Medical Sciences (CIOMS) Working Group. Drug-induced liver injury (DILI): Current status and future directions for drug development and the post-market setting. 2020. <https://cioms.ch/publications/product/drug-induced-liver-injury>. Accessed 31 May 2021.
2. García-Cortés M, Ortega-Alonso A, Lucena MI, Andrade RJ, Grupo de Estudio para las Hepatopatías Asociadas a Medicamentos (GEHAM). Drug-induced liver injury: a safety review. *Expert Opin Drug Saf.* 2018;17(8):795-804. <https://doi.org/10.1080/14740338.2018.1505861>
3. Stravitz RT, Lee WM. Acute liver failure. *Lancet.* 2019;394(10201):869–81. [https://doi.org/10.1016/S0140-6736\(19\)31894-X](https://doi.org/10.1016/S0140-6736(19)31894-X).
4. Stevens JL, Baker TK. The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov Today.* 2009;14(3–4):162–7. <https://doi.org/10.1016/j.drudis.2008.11.009>.
5. Issa AM, Phillips KA, Van Bebber S, et al. Drug withdrawals in the United States: a systematic review of the evidence and analysis of trends. *Curr Drug Saf.* 2007;2(3):177–85. <https://doi.org/10.2174/157488607781668855>.

6. Solotke MT, Dhruva SS, Downing NS, Shah ND, Ross JS. New and incremental FDA black box warnings from 2008 to 2015. *Expert Opin Drug Saf.* 2018;17(2):117–23. <https://doi.org/10.1080/14740338.2018.1415323>.
7. Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med.* 2016;14:10. <https://doi.org/10.1186/s12916-016-0553-2>.
8. Wysowski DK, Swartz L. Adverse drug event surveillance and drug withdrawals in the United States, 1969–2002: the importance of reporting suspected reactions. *Arch Intern Med.* 2005;165(12):1363–9. <https://doi.org/10.1001/archinte.165.12.1363>.
9. García-Cortés M, Robles-Díaz M, Stephens C, Ortega-Alonso A, Lucena MI, Andrade RJ. Drug induced liver injury: an update. *Arch Toxicol.* 2020;94(10):3381–407. <https://doi.org/10.1007/s00204-020-02885-1>.
10. McGill MR, Jaeschke H. Biomarkers of drug-induced liver injury. *Adv Pharmacol.* 2019;85:221–39. <https://doi.org/10.1016/bs.apha.2019.02.001>.
11. Real M, Barnhill MS, Higley C, Rosenberg J, Lewis JH. Drug-induced liver injury: highlights of the recent literature. *Drug Saf.* 2019;42(3):365–97. <https://doi.org/10.1007/s40264-018-0743-2>.
12. United States Department of Health and Human Services Food and Drug Administration (FDA) Center for Drug Evaluation and Research, FDA Center for Biologics Evaluation and Research. Drug-induced liver injury: premarketing clinical evaluation. 2009. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/drug-induced-liver-injury-premarketing-clinical-evaluation>. Accessed 31 May 2021.
13. Aithal GP, Watkins PB, Andrade RJ, et al. Case definition and phenotype standardization in drug-induced liver injury. *Clin Pharmacol Ther.* 2011;89(6):806–15. <https://doi.org/10.1038/clpt.2011.58>.
14. Andrade RJ, Chalasani N, Björnsson ES, et al. Drug-induced liver injury. *Nat Rev Dis Primers.* 2019;5(1):58. <https://doi.org/10.1038/s41572-019-0105-0>.
15. Danan G, Teschke R. Drug-induced liver injury: why is the Roussel Uclaf Causality Assessment Method (RUCAM) still used 25 years after its launch? *Drug Saf.* 2018;41(8):735–43. <https://doi.org/10.1007/s40264-018-0654-2>.
16. Hayashi PH, Barnhart HX, Fontana RJ, et al. Reliability of causality assessment for drug, herbal and dietary supplement hepatotoxicity in the Drug-Induced Liver Injury Network (DILIN). *Liver Int.* 2015;35(5):1623–32. <https://doi.org/10.1111/liv.12540>.
17. Tan EH, Low EXS, Dan YY, Tai BC. Systematic review and meta-analysis of algorithms used to identify drug-induced liver injury (DILI) in health record databases. *Liver Int.* 2017;38(4):742–53. <https://doi.org/10.1111/liv.13646>.
18. Council for International Organizations of Medical Sciences (CIOMS) Working Group. Development and rational use of standardised MedDRA queries (SMQs): retrieving adverse drug reactions with MedDRA. 2nd ed. Geneva: Council for International Organizations of Medical Sciences (CIOMS); 2016.
19. Tansel A, Kanwal F, Hollinger FB. Use of Hy’s Law, R criteria, and nR criteria to predict acute liver failure or transplantation in patients with drug-induced liver injury. *Gastroenterology.* 2015;148(2):452. <https://doi.org/10.1053/j.gastro.2014.11.046>.
20. Rosner B. *Fundamentals of biostatistics.* 4th ed. Boston: Cengage Learning; 2015.
21. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology.* 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
22. Gordis L. *Epidemiology.* 5th ed. Philadelphia: Saunders; 2014.
23. Tarantino G, Di Minno MN, Capone D. Drug-induced liver injury: is it somehow foreseeable? *World J Gastroenterol.* 2009;15(23):2817–33. <https://doi.org/10.3748/wjg.15.2817>.
24. Tischer S, Fontana RJ. Drug-drug interactions with oral anti-HCV agents and idiosyncratic hepatotoxicity in the liver transplant setting. *J Hepatol.* 2014;60(4):872–84. <https://doi.org/10.1016/j.jhep.2013.11.013>.
25. Daly AK. Are polymorphisms in genes relevant to drug disposition predictors of susceptibility to drug-induced liver injury? *Pharm Res.* 2017;34(8):1564–9. <https://doi.org/10.1007/s11095-016-2091-1>.
26. Tarantino G, Conca P, Basile V, et al. A prospective study of acute drug-induced liver injury in patients suffering from non-alcoholic fatty liver disease. *Hepatol Res.* 2007;37(6):410–5. <https://doi.org/10.1111/j.1872-034X.2007.00072.x>.
27. Stine JG, Chalasani NP. Drug hepatotoxicity: environmental factors. *Clin Liver Dis.* 2017;21(1):103–13. <https://doi.org/10.1016/j.cld.2016.08.008>.

28. González-Jimenez A, McEuen K, Chen M, et al. The influence of drug properties and host factors on delayed onset of symptoms in drug-induced liver injury. *Liver Int.* 2019;39(2):401–10. <https://doi.org/10.1111/liv.13952>.
29. Tan EH, Ling ZJ, Ang PS, Sung C, Dan YY, Tai BC. Comparison of laboratory threshold criteria in drug-induced liver injury detection algorithms for use in pharmacovigilance. *Pharmacoepidemiol Drug Saf.* 2020. <https://doi.org/10.1002/pds.5099>.
30. Basile AO, Yahi A, Tatonetti NP. Artificial intelligence for drug toxicity and safety. *Trends Pharmacol Sci.* 2019;40(9):624–35. <https://doi.org/10.1016/j.tips.2019.07.005>.
31. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc.* 2019;26(4):364–79. <https://doi.org/10.1093/jamia/ocy173>.
32. Minerali E, Foil DH, Zorn KM, Lane TR, Ekins S. Comparing machine learning algorithms for predicting drug-induced liver injury (DILI). *Mol Pharm.* 2020;17(7):2628–37. <https://doi.org/10.1021/acs.molpharmaceut.0c00326>.
33. Teschke R. DILI, HILI, RUCAM algorithm, and AI, the artificial intelligence: provocative issues, progress, and proposals. *Arch Gastroenterol Res.* 2020;1(1):4–11. <https://doi.org/10.33696/Gastroenterology.1.002>.
34. Sassolas B, Haddad C, Mockenhaupt M, et al. ALDEN, an algorithm for assessment of drug causality in Stevens-Johnson syndrome and toxic epidermal necrolysis: comparison with case-control analysis. *Clin Pharmacol Ther.* 2010;88(1):60–8. <https://doi.org/10.1038/clpt.2009.252>.
35. Mato AR, Riccio BE, Qin L, et al. A predictive model for the detection of tumor lysis syndrome during AML induction therapy. *Leuk Lymph.* 2006;47(5):877–83. <https://doi.org/10.1080/10428190500404662>.
36. Perry PJ, Wilborn CA. Serotonin syndrome vs neuroleptic malignant syndrome: a contrast of causes, diagnoses, and management. *Ann Clin Psychiatry.* 2012;24(2):155–62.
37. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med.* 2018;46(7):1070–7. <https://doi.org/10.1097/CCM.0000000000003123>.