

An Application of Machine Learning in Pharmacovigilance: Estimating Likely Patient Genotype From Phenotypical Manifestations of Fluoropyrimidine Toxicity

Luis Correia Pinheiro^{1,*}, Julie Durand¹ and Jean-Michel Dogné^{2,3}

Dihydropyrimidine dehydrogenase (DPD)-deficient patients might only become aware of their genotype after exposure to dihydropyrimidines, if testing is performed. Case reports to pharmacovigilance databases might only contain phenotypical manifestations of DPD, without information on the genotype. This poses a difficulty in estimating the cases due to DPD. Auto machine learning models were developed to train patterns of phenotypical manifestations of toxicity, which were then used as a surrogate to estimate the number of cases of DPD-related toxicity. Results indicate that between 8,878 (7.0%) and 16,549 (13.1%) patients have a profile similar to DPD deficient status. Results of the analysis of variable importance match the known end-organ damage of DPD-related toxicity, however, accuracies in the range of 90% suggest presence of overfitting, thus, results need to be interpreted carefully. This study shows the potential for use of machine learning in the regulatory context but additional studies are required to better understand regulatory applicability.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

✓ The true proportion of cases of dihydropyrimidine dehydrogenase (DPD)-deficient individuals with adverse reactions to fluoropyrimidines reported to pharmacovigilance databases is not estimable using traditional methods.

WHAT QUESTION DID THIS STUDY ADDRESS?

✓ This study aimed at providing an estimate of the proportion of patients that might have susceptibility to fluoropyrimidine toxicity due to DPD deficiency. In essence, this is an imputation exercise, using machine learning models to classify likely DPD genotype.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

✓ Machine learning models can assist in imputing likely genotype from phenotypical manifestations. The results allow a better understanding of the influence of DPD deficiency in reports of adverse drug reactions.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

✓ Machine learning models applied to large pharmacovigilance databases can help answer certain research questions, which were difficult to address with more traditional methods.

Fluorouracil (5-FU) is a fluoropyrimidine anticancer drug that has been used in the treatment of solid tumors for decades.¹ About 10% of patients taking 5-FU undergoes renal excretion, whereas over 80% are cleared by catabolic degradation. Dihydropyrimidine dehydrogenase (DPD) is the initial and rate-limiting enzyme in the catabolism of 5-FU.² Many variants of the *DPYD* gene, which encodes for DPD, have been described, however, only a few have been shown to lead to absent or reduced enzyme activity.³

Although most individuals with partial DPD deficiency unexposed to fluoropyrimidines do not exhibit obvious symptoms, infants with severe DPD deficiency may have neurological problems, such as recurrent seizures, intellectual disability, microcephaly,

hypertonia, and autistic behaviors, among others. All patients with DPD deficiency, regardless of whether they show any symptoms, are susceptible to serious, sometimes fatal adverse reactions, on exposure to fluoropyrimidines. Fluoropyrimidine toxicity in partial DPD-deficient individuals may manifest as severe mucositis, neutropenia, thrombocytopenia, hemorrhage, hand-foot syndrome, diarrhea, dyspnea, and alopecia.^{4,5} The prevalence of DPD deficiency seems to be dependent on race and sex. African American women showed the highest prevalence of DPD deficiency compared with African American men, white women, and white men (12.3%, 4.0%, 3.5%, and 1.9%, respectively).⁶

Low or absent DPD activity can lead to fluoropyrimidine-associated toxicity, which occurs in about 30% of treated patients

¹European Medicines Agency, Amsterdam, The Netherlands; ²University of Namur, Namur, Belgium; ³Belgian Federal Agency for Medicines and Health Products, Brussels, Belgium. *Correspondence: Luis Correia Pinheiro (luis.pinheiro@ema.europa.eu)

Received September 16, 2019; accepted January 3, 2020. doi:10.1002/cpt.1789

with 0.5–1% having fatal treatment-related toxicity.^{7,8} However, adverse drug reactions (ADRs) to 5-FU and related substances (capecitabine, tegafur, and flucytosine) can also occur in non-DPD-deficient patients.

In March 2019, the French regulatory authority (ANSM) notified the European Medicines Agency's (EMA) Pharmacovigilance Risk Assessment Committee of a referral under Article 31 of Directive 2001/83/EC.⁹ The referral involved a review of available data on screening methods to detect DPD deficiency with the aim of recommending changes to ensure the safe use of 5-FU and related drugs.

In the context of this referral, an analysis of data in the EudraVigilance (EV) database was performed to estimate the number (or proportion) of individual case safety reports (ICSRs) to fluorouracil and related substances that might have been due to DPD deficiency.

EV is the system for collecting, managing, and analyzing suspected ADRs to medicines authorized in the European Economic Area (EEA). By the end of 2017, the EV held a total of 7,948,873 individual cases. Of the cases submitted in a postauthorization setting, 64% were submitted from outside the EEA and 36% from EEA countries. Thus, the EV provides a rich dataset with global representation.¹⁰

The traditional approach to estimating counts of an ADR is to develop a case definition using the Medical Dictionary for Regulatory Activities (MedDRA).¹¹ MedDRA (version 21.1) has a preferred term for “Dihydropyrimidine dehydrogenase deficiency.” However, it is anticipated that it would only be used in patients with known DPD deficiency at the time of reporting, which is not always the case. A standardized MedDRA query would have been helpful to extract information and act as a *de facto* case definition,¹² however, there is none for DPD-related toxicities.

A potential alternative to developing a time-consuming consensus-based case definition is to use machine learning models. Supervised classification machine learning models can help identify complex relationships in the variables (or features) and assign a probabilistic classification to each observation.¹³

By using the phenotypical manifestations (i.e., the adverse events associated with 5-FU and derivatives) we hypothesize that we might be able to classify the likely genotype of the patient. To the best of our knowledge, this is a novel approach to classifying cases in pharmacovigilance databases, but the use of novel methods to build bespoke case definitions has been attempted before.¹⁴

In this study, multivariable prediction models were developed and validated. The resulting models were applied to case reports of 5-FU and derivatives. With all patients classified it became possible to estimate the number of ADRs where the patients are likely DPD deficient.

METHODS

Data

The data were sourced from the European Union's central database of reports of suspected ADRs, EV. The period of interest was from the start of data collection to March 15, 2019.

Exposure

The exposure was defined as use of fluorouracil and fluorouracil-related substances, namely capecitabine, fluorouracil, tegafur, and flucytosine containing medicinal products.

Classification in the training set

This study was a two-class classification problem, where cases were classified as likely DPD deficient (true positives) and likely not DPD deficient (true negatives). Cases were classified as DPD deficient (true positives) where the MedDRA term “dihydropyrimidine dehydrogenase deficiency” was reported as a reaction, or medical history, or a test result was reported, suggestive of DPD deficiency.

There were very few cases of individuals where normal DPD activity is reported as such in the database (i.e., DPD laboratory testing reported in the case did not suggest deficiency). Thus, likely non-DPD-deficient cases were defined as case reports of medicinal products used in similar indications as fluorouracil and fluorouracil-related substances, but which do not have a DPD interaction. Trastuzumab, pembrolizumab, docetaxel, and irinotecan were used. Non-DPD-deficient cases were randomly selected on a 1:4 relation. This means that for each DPD-deficient case, four non-DPD-deficient cases were assigned.

Non-DPD-deficient cases that also reported exposure to fluorouracil or related products were removed from the sample, except if they were known to be DPD deficient (i.e., it was reported in the case).

Considering that the sample defined the distribution of the classes is imbalanced (i.e., 20% DPD-deficient and 80% non-DPD-deficient), the classes were balanced in the machine learning model. The auto machine learning model does this algorithmically, and includes a blend of under-sampling the majority class and resampling the minority class.

Training, validation, and testing set

The data were partitioned in a 75% training set and a 25% test set. Five-fold cross-validation was performed in the training set.

Features

The features (or variables) used were age, sex, and adverse reaction reported at high level term. The high level term with the preferred term “Dihydropyrimidine dehydrogenase deficiency” was excluded as it is highly correlated with DPD deficiency.

Analytics

Auto machine learning was applied using H₂O for R. Distributed random forests, gradient boosting machines, and generalized linear models (GLMs) were included. H₂O is an open source lightweight in-memory machine learning platform written in Java.¹⁵

RESULTS

As of March 15, 2019, there were a total of 126,890 ICSRs in EV with capecitabine, fluorouracil, tegafur, or flucytosine containing medicinal products reported as suspect, interacting, or concomitant. DPD deficiency could be ascertained in 260 cases: 184 had information in the medical history, 76 in the reaction field, and 3 in the laboratory tests fields (some cases fulfilled more than one criterion).

The models were trained in a data frame containing 1,185 observations and 434 features. The five most accurate trained models had an accuracy that ranged between 0.8855 and 0.9192. The model with the best accuracy had a precision of 0.9331 and a specificity of 0.6792 (Table 1).

When the models were applied to the 126,630 cases without established DPD deficiency, the estimate of cases possibly due to DPD deficiency ranged from 8,878 (7.0%) to 16,549 (13.1%) (Table 2).

Gradient boosting machine (GBM) models, contrary to ensemble or GLMs, provide information on the relative importance of the features. Eight of the top 10 most important features in the 2 GBM models are the same (Table 3).

Table 1 Performance metrics, at a probabilistic threshold of 50%, for the validation (i.e., test set) for the top five models, ranked by accuracy

Model identification	Accuracy	Precision	Recall	Specificity
Stacked ensemble best of family	0.9192	0.9331	0.9713	0.6792
Stacked ensemble all models	0.9125	0.9325	0.9631	0.6792
GBM grid 1 model 12	0.8990	0.9385	0.9385	0.7170
GBM grid 1 model 13	0.8956	0.9019	0.9795	0.5094
GLM grid 1 model 1	0.8855	0.9038	0.9631	0.5283

GBM, gradient boosting machine; GLM, generalized linear model.

Table 2 Estimate of the number of cases likely to be related to DPD deficiency, for the top five models, ranked by model accuracy

Model identification	Estimate of cases N (%)
Stacked ensemble all models	8,878 (7.0)
Stacked ensemble best of family	16,549 (13.1)
GBM grid 1 model 12	14,604 (11.5)
GBM grid 1 model 13	10,944 (8.6)
GLM grid 1 model 1	11,432 (9.0)

DPD, dihydropyrimidine dehydrogenase; GBM, gradient boosting machine; GLM, generalized linear model.

Table 3 Variable importance for the GBM models, showing top 10 most important features

Model identification	Features
GBM grid 1 model 12	HLT Mucosal findings abnormal Age HLT Poisoning and toxicity HLT Marrow depression and hypoplastic anemias HLT Stomatitis and ulceration Sex HLT Thrombocytopenias HLT Diarrhea (excluding infective) HLT Neutropenias HLT Dermatitis ascribed to specific agent
GBM grid 1 model 13	HLT Mucosal findings abnormal HLT Poisoning and toxicity HLT Diarrhea (excluding infective) Age HLT Marrow depression and hypoplastic anemias HLT Leukopenias HLT Thrombocytopenias HLT Dermatitis ascribed to specific agent HLT Stomatitis and ulceration HLT Neutropenias

GBM, gradient boosting machine; HLT, high level term.

DISCUSSION

The auto machine learning models suggest that between 7% and 13% of the cases of toxicity with 5-FU and derivatives in the EV database might be due to DPD deficiency. This estimate is very close to the 2–12% estimate of the general population that may be vulnerable to toxic reactions to fluoropyrimidine drugs due to DPD deficiency⁴.

Furthermore, the features of highest importance, as highlighted by the gradient boosting machine models, are closely related to the typical manifestations of DPD-related toxicity, namely mucosal problems, diarrhea, thrombocytopenia, and skin reactions. Furthermore, sex appears in one model as an important feature, which is in line with the known sex imbalance in DPD deficiency (i.e., female sex seems to have higher prevalence of DPD deficiency).

This result could be read as the models identifying correctly the main manifestations of DPD-related toxicity. It is reassuring regarding the wider validity of the models, as the important features of the models are not at odds with the known pattern for the disease.

Of note, descriptive analyses run in parallel using an *ad hoc* case definition aimed at capturing all reactions within the spectrum of possible DPD-related toxicity yielded an estimate of 33.9% of cases (data not shown).

As this was a novel application of machine learning models to reach an overall estimate of counts of cases, the use of auto machine learning models, whereby the iterative model building and hyperparameter tuning is automated—was considered sufficient. The best performing models were stacked ensemble models (i.e., models based on a combination of machine learning models). These are known to have better predictive abilities than individual models and the results support that. However, these models come at a cost of interpretation; GBM models provide information on variable importance and GLMs provide the variable coefficients but stacked ensembles do not provide easily interpretable information on variable importance. In this study, having insights into variable importance is useful in understanding the validity of the models.

Machine learning models learn patterns (or correlations), therefore, it is not adequate to assume causal relationships from these results. Models identify patterns even in the sample error space—also known as overfitting. The feature age, for instance, may solely indicate a higher risk of having a malignant disease or general pharmacokinetic changes and not a specific DPD relationship.

Furthermore, considering the simplicity of the features used and complexity of the prediction problem, the performance metrics, with accuracies around 90%, might suggest some overfitting. To improve validity of the models, a set of oncology drugs—two biologicals and two small molecule products—which do not have a DPD metabolic pathway were chosen. It is possible that the safety profiles of these are so distinct from that of fluoropyrimidines that it leads to overfitting.

In addition, although a selection of a 1:4 ratio of DPD deficient to non-DPD deficient case reports was made, the models chosen

balanced the classes by oversampling from the minority class and undersampling from the majority class.

Estimating the number of patients likely to be DPD deficient but otherwise asymptomatic is a particularly difficult topic regardless of the methodology chosen. The pathological mechanism is through reduced clearance of fluoropyrimidines and, thus, manifestations are similar in DPD-deficient and nondeficient patients. In addition, they also overlap with the adverse reaction profile of other oncologic products.

The hypothesis for this study was that a pattern or combination of manifestations could potentially be characteristic of DPD-related toxicity. A cautious interpretation is that machine learning models indicate that 7–13% of all case reports to fluoropyrimidines have an adverse reaction profile similar to that reported in DPD-deficient patients. This frequency is within the range of prevalence of DPD deficiency and the variable importance follows the expected profile of the reaction and sex imbalances.

From a regulatory point of view, this study provides a good example in one of the key areas of the EMA's regulatory strategy of exploiting artificial intelligence in regulatory decision making¹⁶ and is the first time that machine learning is used as a pharmacovigilance approach in estimating the number and proportion of ICSRs with a drug that might have been due to a genotype deficiency, notably in the absence of information on the DPD status in the narrative. This will surely help the scientific assessors in estimating the level of the risk and identify proportional risk minimizations strategies of toxicity with 5-FU and derivatives. It is anticipated that such a method will be validated with other examples in the EV database and may be used to provide rapid and reliable estimates as part of other safety procedures, including signal management and continuous monitoring of drug safety.

The European regulatory system is scientifically robust, and coordination of pharmacovigilance activities by the EMA is determinant for safety monitoring of medicines across Europe. While existing methods are well established, this study illustrates how, taking into account their limitations, the use of machine learning models for case definition in pharmacovigilance may provide additional insights that strengthen the evidence base for decision making.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

All authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

All authors wrote the manuscript. L.C.P. designed the research. L.C.P. performed the research. All authors analyzed the data.

DISCLAIMER

The views expressed in this paper are the personal views of the authors and may not be understood or quoted as being made on behalf of or

reflecting the position of the agencies or organizations with which the authors are affiliated.

© 2020 European Medicines Agency. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Wilson, W.L. Chemotherapy of human solid tumors with 5-fluorouracil. *Cancer* **13**, 1230–1239 (1960).
2. Iyer, L. & Ratain, M.J. 5-fluorouracil pharmacokinetics: causes for variability and strategies for modulation in cancer chemotherapy. *Cancer Invest.* **17**, 494–506 (1999).
3. Murphy, C. *et al.* Cost implications of reactive versus prospective testing for dihydropyrimidine dehydrogenase deficiency in patients with colorectal cancer: a single-institution experience. *Dose Response* **16**, 1–6 (2018).
4. Ezzeldin, H. & Diasio, R. Dihydropyrimidine dehydrogenase deficiency, a pharmacogenetic syndrome associated with potentially life-threatening toxicity following 5-fluorouracil administration. *Clin. Colorectal Cancer* **4**, 181–189 (2004).
5. Genetic and Rare Diseases Information Center Dihydropyrimidine dehydrogenase deficiency <<https://rarediseases.info.nih.gov/diseases/19/dihydropyrimidine-dehydrogenase-deficiency>>. Accessed September 15, 2019.
6. Mattison, L.K. *et al.* Increased prevalence of dihydropyrimidine dehydrogenase deficiency in African-Americans compared with Caucasians. *Clin. Cancer Res.* **12**, 5491–5495 (2006).
7. van Kuilenburg, A.B. Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. *Eur. J. Cancer* **40**, 939–950 (2004).
8. Mikhail, S.E., Sun, J.F. & Marshall, J.L. Safety of capecitabine: a review. *Expert Opin. Drug Saf.* **9**, 831–841 (2010).
9. European Medicines Agency Fluorouracil and fluorouracil related substances (capecitabine, tegafur and flucytosine) containing medicinal products <<https://www.ema.europa.eu/en/medicines/human/referrals/fluorouracil-fluorouracil-related-substances-capecitabine-tegafur-flucytosine-containing-medicinal>> (2019). Accessed September 15, 2019.
10. Postigo, R. *et al.* EudraVigilance medicines safety database: publicly accessible data for research and public health protection. *Drug Saf.* **41**, 665–675 (2018).
11. Meddra.org MedDRA|Medical Dictionary for Regulatory Activities <<http://www.meddra.org/>> (2015). Accessed September 15, 2019.
12. Mozzicato, P. Standardised MedDRA queries. *Drug Saf.* **30**, 617–619 (2007).
13. Sidey-Gibbons, J.A.M. & Sidey-Gibbons, C.J. Machine learning in medicine: a practical introduction. *BMC Med. Res. Methodol.* **19**, 64 (2019).
14. Lethebe, B.C. *et al.* Developing a case definition for type 1 diabetes mellitus in a primary care electronic medical record database: an exploratory study. *C. Open* **7**, E246–E251 (2019).
15. H₂O. Open source leader in AI and ML <<https://www.h2o.ai/>>. Accessed September 15, 2019.
16. European Medicines Agency. EMA regulatory science to 2025—strategic reflection <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatoryscience-2025-strategic-reflection_en.pdf> (2018).