

SCIENTIFIC REPORTS



OPEN

Predicting the need for a reduced drug dose, at first prescription

Adrien Coulet^{1,2}, Nigam H. Shah², Maxime Wack³, Mohammad B. Chawki³, Nicolas Jay^{1,3} & Michel Dumontier^{2,4}

Prescribing the right drug with the right dose is a central tenet of precision medicine. We examined the use of patients' prior Electronic Health Records to predict a reduction in drug dosage. We focus on drugs that interact with the P450 enzyme family, because their dosage is known to be sensitive and variable. We extracted diagnostic codes, conditions reported in clinical notes, and laboratory orders from Stanford's clinical data warehouse to construct cohorts of patients that either did or did not need a dose change. After feature selection, we trained models to predict the patients who will (or will not) require a dose change after being prescribed one of 34 drugs across 23 drug classes. Overall, we can predict ($AUC \geq 0.70$ – 0.95) a dose reduction for 23 drugs and 22 drug classes. Several of these drugs are associated with clinical guidelines that recommend dose reduction exclusively in the case of adverse reaction. For these cases, a reduction in dosage may be considered as a surrogate for an adverse reaction, which our system could indirectly help predict and prevent. Our study illustrates the role machine learning may take in providing guidance in setting the starting dose for drugs associated with response variability.

Precision medicine aims to improve clinical care using an individual's information such as genetics, lifestyle, and environment¹. Considering such information may help in prescribing the right drug at the right dose, and in the process reduce adverse drug reactions (ADR), which are estimated to account for one-third of hospital adverse events and approximately 280,000 hospital admission annually in the US^{2,3}. The inter-individual variability in drug responses, including ADR, may have diverse causes such as: *patient conditions*, e.g. a renal dysfunction impacts response to renally excreted drugs; *drug interactions*, e.g. a first drug, such as *fluoxetine* may inhibit the effect of a second drug such as *tamoxifen* by targeting the same enzyme; *drug-food interactions* e.g. aliments such as grapefruit may inhibit drug metabolism enzymes, such as CYP3A4, causing drug toxicity; *genetics* e.g. a genomic variation in the coding sequence of a drug metabolizing enzyme such as CYP3A4 also impacts drug response. The variety of factors, both known and suspected or unknown, makes it challenging for health institutions to take proper precautions⁴.

Electronic Health Records (EHRs) offer novel opportunities for using patient data to study variable patient outcomes including drug response⁵. Research has explored the secondary use of EHRs for predicting disease occurrence⁶, drug effects and their interactions⁷, detecting higher rates of adverse events⁸, and identifying subgroups of drug responses⁹. PheWAS (Phenotype Wide Association Studies) explore the association of a genetic variation with multiple disease phenotypes¹⁰. Several efforts have associated genotyping from biobanks with clinical data in EHRs^{11–13} and evaluated their potential importance¹⁴. However, given that genotype data is not yet routinely available in EHRs, an alternative is to use the recorded phenotypes as surrogate markers of individual variations that lead to differential drug response. To the best of our knowledge, no work has yet focused on using EHR data, in absence of genetic information, to predict the variable response to drug exposure.

The goal of this research is to examine the feasibility of using phenotypic data of an individual, recorded in their EHR prior to the drug exposure, to predict a reduced drug-dosing event. Our approach begins with a feature selection step to identify patient characteristics (phenotypes) that are over-represented in patients that needed a drug dose change (increase or reduction) *vs.* those who did not. These selected features comprise *phenotype profiles* that are used for data reduction, prediction and interpretation.

In our work, we considered *drug dose changes* as a sign for individual variation in drug response. In such cases, one may consider the reduction in dosage of a drug as the result of a potential adverse reaction, while an increase

¹Université de Lorraine, CNRS, Inria, LORIA, 54000, Nancy, France. ²Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, California, USA. ³Service d'Evaluation et d'Information Médicales, University Hospital of Nancy (CHRU), Nancy, France. ⁴Institute of Data Science, Maastricht University, Maastricht, Netherlands. Correspondence and requests for materials should be addressed to A.C. (email: adrien.coulet@loria.fr)

Received: 4 October 2017

Accepted: 6 October 2018

Published online: 22 October 2018

in dosage suggests a lack of response. If, however, the dose is unchanged, it suggests an appropriate dosing for the patient.

We focus our study on a set of drugs known to be associated with high inter-individual variability in response, i.e. drugs metabolized by enzymes of the P450 cytochrome family¹⁵. Genes that code for P450 enzymes have been extensively studied in pharmacogenomics because their variations impact the activity of P450 enzyme and in turn, drug metabolism.

Results

We examined 34 drugs that are metabolized by P450 enzymes, for which we were able to observe at least 300 drug prescription intervals as described in the methods. We constructed phenotype profiles from patients prescribed one of the 34 drugs by comparing those who experienced a drug dose reduction with those whose dose had not been changed (dose continuation) and by comparing patients who experienced a drug dose increase with those who had a dose continuation. We subsequently evaluated whether phenotype profile were effective in predicting patient drug sensitivity. We found that this method can successfully predict dose reductions for most drugs (23 out of 34), but it could not predict dose increases. We developed a web interface displaying phenotype profiles and the predictions for dose changes for our drugs and their subgroups. This tool was used by three physicians (MW, MC and NJ) to provide clinical interpretation of the reasons why these features may play a role in predicting the need for a lower dose.

Constructing phenotype profiles for drug dose change. Each profile is composed of three *types* of features: diagnostic codes, conditions mentioned in clinical notes and laboratory test orders. Figure 1 shows an example of phenotype profile, containing the top 10 features of each type that are enriched in patients who needed a dose reduction for *tacrolimus*, an immunosuppressant. We generated two types phenotype profiles: the first contains 300 features, composed of the top 100 features of each type (diagnostic codes, conditions, lab orders) and the second is composed of all the features that meet statistical significance. For some drugs, no feature meets with statistical significance, leading to an empty phenotype profile—which means there are no specific characteristics in the patient record that could be indicative of needing a dose change. To allow interpretation of our results, we built a *phenotype profile browser*, available at <http://snowflake.loria.fr/p450/>. It enables browsing phenotype profiles of drugs and drug sets, and displays the top features of each type. To aid interpretation, phenotype profiles also show the number of publications found in PubMed that mention both the drug and the code/condition/lab test included in the profile.

Predicting drug dose reduction. We evaluated the effectiveness of phenotype profiles to predict the need for a patient to receive a dose-reduction to a single drug (or to a set of drugs) in two ways: first, by 10-fold cross-validation and second, by holding out the last year of data as a test-set. The first evaluation is a standard setup in machine learning. For the second, the prediction model was trained on patient data from 2008–2013, then tested on the last year of data, i.e., 2014. Drugs and drug sets with less than 300 instances or with an empty phenotype profile were excluded from the evaluation; leaving 34 drugs and 23 drug sets.

Evaluation using 10-fold cross-validation yielded an average AUC-ROC of 0.76 and F-measure of 0.69. Average AUC-ROC is 0.68, and F-measure is 0.64, in the hold out evaluation. Results of the dose reduction prediction with an F-measure greater than or equal to 0.7 in the held out evaluation are reported in Table 1. Seven of the top ten results of both evaluations are identical. This difference in discrimination accuracy is partially due to the different sizes of training sets, which is smaller for the second evaluation set up. For single drugs or sets of drugs, both evaluations obtain high values in some cases. For instance, with the 10-fold cross validation we obtain 0.94 AUC and 0.89 F-measure for tacrolimus; 0.93 AUC and 0.88 F-measure for the ATC class L-Antineoplastic and Immunomodulating agents. 23 and 10 drugs obtained an AUC \geq 0.70 with the 10-fold cross validation and the hold out validation, respectively.

Table 2 shows performances using different feature sets for the drug dose reduction prediction. Using all the features does not result in significantly different performance (p-value > 0.05, t-test) than using only 300 features (i.e., the top 100 features of each type). Average F-measures obtained for diagnostic codes, conditions, and lab orders are 0.37, 0.41 and 0.69, respectively. For most of the drugs and drug sets, lab orders constitutes the most important type of features, while diagnostic codes contributes the least. We are unable to predict dose increases using phenotype profiles (AUC-ROC of 0.53 and 0.40 for the hold last year out and the 10-fold cross-validation).

Interpretation. We reviewed prescription guidelines and sections “*dosage and administration*” of drug labels for the 23 drugs for which we were able to predict dose reduction to verify if dose reductions were recommended in the case of an undesirable response, or if it may be part of a regular protocol of drug prescription. For 14 drugs, out of 23, a dose reduction is only recommended in the case of an undesirable response; For the other 8 drugs, a dose reduction is recommended either in the case of an undesirable response or in normal prescription management. In our top-prediction list provided Table 2, dose reduction is recommended only in the case of an undesirable response in 5 out of 7 drug guidelines; dose reduction is part of the normal prescription for the other 2 drugs (hydrocortisone and methadone). Lists of drugs and their guidelines are provided in Supplementary file S1.

We examined phenotype profiles of four drugs (labetalol, tacrolimus, warfarin and sildenafil) to understand why phenotype profiles enable the prediction of dose reductions for those drugs before their first prescription. Phenotype profiles reviewed are in Fig. 1 and in Supplementary Fig. S2.

The profile of tacrolimus, an immunosuppressant used in patients with organ transplants, shows that dose reduction intervals for this drug (in comparison with drug continuation intervals) is negatively associated with previous urine drug analysis (see Fig. 1). We also found that on average there were more intervals of dose reduction per patient for this drug than for others (i.e., 3.13 dose reductions per patient for tacrolimus and 1.44

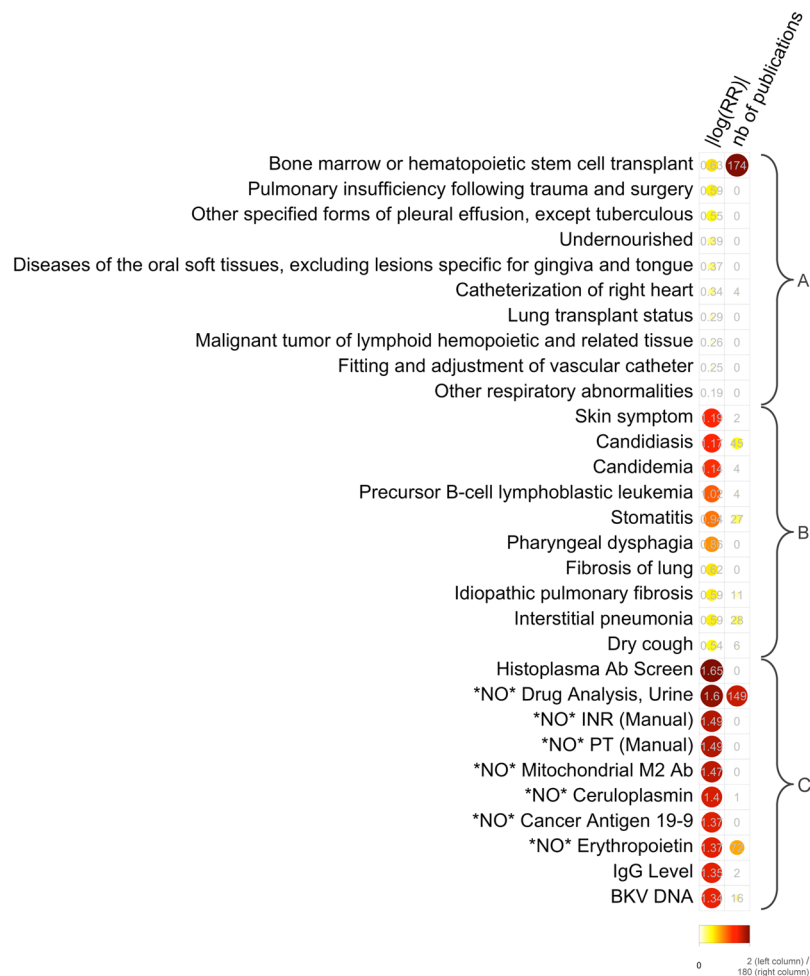


Figure 1. Example of phenotype profile. The shown phenotype profile comprises of the top 10 diagnoses (A), condition-mentions (B), and lab tests (C) that are seen before a tacrolimus prescription in patients who subsequently needed a dose reduction. Only the top 10 in each type are shown. Each phenotype is associated with a statistically significant p-value (hypergeometric test, $p < 0.05$, Bonferroni correction for multiple testing), and ordered by the absolute value of the log of the Risk Ratio (on a 0 to 2 scale), in the first column. For interpretation purpose, the number of articles in PubMed that mention both the drug and the phenotype is also provided, in the second column (on a 0 to 180 scale). For example, 45 articles mention both *tacrolimus* and *candidiasis*. (A) Diagnoses are ICD-9-CM codes associated with patient visits; (B) Conditions are phenotypic terms mentioned in text of clinical notes; (C) Lab tests are orders of laboratory tests. Lab codes prefixed with the string “*NO*” indicate a negative relationship between the lab test and the drug dose decrease.

(SD = 0.40) dose reductions on average for all drugs). This, and association with urine drug analysis, can be explained by the fact that tacrolimus dosage is often re-evaluated to get to the optimal dosage, with the help of urine drug analysis. The negative association with urine analysis is expected due to more dose continuation following a normal analysis.

The profile of labetalol, an alpha/beta adrenergic antagonist use to treat essential hypertension, shows that its dose reduction is strongly and positively associated with prerenal renal failure and acute renal failure diagnostics as well as lab tests for renin activity and aldosterone (see Supplementary Fig. S2). These diagnostics are associated with non-essential hypertension, and can explain a dose reduction of the first intention treatment, after specific tests for secondary hypertension. Renin activity and aldosterone testing are indeed used to diagnose secondary causes of hypertension.

Warfarin is an anti-vitamin K anticoagulant used to prevent the formation of blood clots. Interestingly, dose reduction is associated with candidiasis (see Supplementary Fig. S2), a diagnosis that leads to prescription of anti-fungal drugs such as miconazole, which are contra-indicated in conjunction with warfarin, because of their inhibitory interaction with cytochrome P450, of which warfarin is a substrate. Warfarin dose reduction is also associated with chronic renal failure, a contra-indication for warfarin prescription, as well as with clot kinetics lab tests, standard control tests for dose adaptation of anticoagulant drugs. We did not find association with INR (International Normalized Ratio) testing, probably because INR is monitored during dose adjustment as well as routinely after a stable therapeutic dose is found.

Drug or drug set	10-fold cross-validation			Hold last year out		
	instances	AUC-ROC	F-m (P; R)	instances	AUC-ROC	F-m (P; R)
<i>Labetalol</i>	353	0.85	0.77 (0.78; 0.77)	314	0.95	0.86 (0.87; 0.86)
<i>Tacrolimus</i>	709	0.94	0.89 (0.89; 0.89)	598	0.94	0.86 (0.86; 0.86)
<i>Itraconazole</i>	292	0.88	0.80 (0.80; 0.80)	460	0.86	0.85 (0.88; 0.86)
<i>Sildenafil</i>	419	0.88	0.80 (0.81; 0.80)	338	0.90	0.80 (0.80; 0.80)
<i>Methadone</i>	941	0.90	0.81 (0.82; 0.81)	866	0.82	0.74 (0.78; 0.75)
<i>Warfarin</i>	2851	0.82	0.74 (0.74; 0.74)	2630	0.79	0.72 (0.72; 0.72)
<i>Hydrocortisone</i>	4853	0.90	0.83 (0.83; 0.83)	4288	0.80	0.71 (0.72; 0.71)
<i>H</i>	17302	0.86	0.78 (0.78; 0.78)	15366	0.76	0.70 (0.70; 0.70)
<i>2C9</i>	21607	0.79	0.71 (0.71; 0.71)	19212	0.73	0.70 (0.70; 0.70)
<i>L</i>	12119	0.93	0.88 (0.88; 0.88)	10354	—	—
<i>All P450-drugs</i>	91267	0.70	0.64 (0.64; 0.64)	80614	—	—

Table 1. Results of the prediction of dose reduction using phenotype profiles. Two evaluations were performed: 10-fold cross-validation and hold last year out. Only the top 100 conditions mentioned in clinical note, top 100 diagnostic codes and top 100 lab codes with a significant p-value (hypergeometric test, $p < 0.05$, Bonferroni correction for multiple testing) are used in the phenotype profiles. Instances are balanced sets of intervals of dose decrease and dose continuation. Drugs or drug sets associated with a F-measure (F-m) ≥ 0.7 during the hold out evaluation are reported, in addition to results associated with the sets of all P450-drugs and drugs of the ATC class L. L is the first level-ATC classes *Antineoplastic and immunomodulating agents* and H is *Systemic hormonal preparations, excluding sex hormones and insulins*. Results for the ATC class L are the best for the 10-fold cross validation, whereas they are not computable for the hold out validation because of empty phenotype profiles in the prediction year. |instances| refers to the number of instances in the training set. Precision (P) and Recall (R) are provided along with the F-measure. Complete results are available in Supplementary material S1. Instance numbers are counted on the train set only and are averaged over the 10 folds of the cross-validation.

Drug or drug set	F-measure			
	Top 100 diagnostics	Top 100 conditions	Top 100 labs	Top 300 features
<i>Labetalol</i>	0.37	0.41	0.81	0.77
<i>Tacrolimus</i>	0.41	0.43	0.90	0.89
<i>Itraconazole</i>	0.36	0.49	0.79	0.80
<i>Sildenafil</i>	0.39	0.42	0.78	0.80
<i>Methadone</i>	0.38	0.43	0.82	0.81
<i>Warfarin</i>	0.38	0.45	0.76	0.74
<i>Hydrocortisone</i>	0.40	0.49	0.82	0.83
<i>H</i>	0.37	0.48	0.78	0.78
<i>2C9</i>	0.36	0.42	0.71	0.71
<i>L</i>	0.41	0.33	0.87	0.88
<i>All P450-drugs</i>	0.36	0.45	0.64	0.64

Table 2. Classifier performance by the type of phenotypic features. Top 100 lists are ordered on the basis of their p-value. Performances are computed using a 10-fold cross-validation. The top 300 features is the combination of the top 100 features of each type (diagnostic codes, conditions mentioned in clinical notes and lab orders).

Lastly, sildenafil is a phosphodiesterase inhibitor used in erectile dysfunction and pulmonary arterial hypertension. Sildenafil profile show that its dose reduction is associated with the presence of thromboembolus, a major contraindication for sildenafil prescription, as well as with bilateral pleural effusion, a symptom of cardiac congestion, and another contraindication for sildenafil prescription (see Fig. S2). Additionally, we found that the sex ratio in the patients who had a dose reduction was 1:1 for men and women, whereas the sex ratio for dose continuation was closer to three men for one woman, indicating that dose reductions mostly happened in patients for whom sildenafil was prescribed for pulmonary arterial hypertension.

Discussion

We trained classifiers to predict dose changes using three types of features from EHR data. Our underlying assumption is that the need for dose changes is marker for an inappropriate drug response, such as a drug adverse response or lack of response. Our classifier (trained on the phenotypic profiles) successfully predicts dose reductions for 23 drugs out of 34 drugs in the study, but is ineffective in predicting dose increases. We also found that the laboratory test orders generated better predictions than other features. Finally, we find that when evaluating

the prediction ability using a held out test set from a later calendar year (i.e., totally unseen ‘future data’), the performance is lower than, but consistent with, the commonly done 10-fold cross validation.

For predictive models in medicine, an unseen *future* dataset is the most appropriate evaluation setup because it mirrors real-life usage of a model. Even in this stringent setting, we are able to predict dose reductions for 10 of the 29 drugs. Note, that in this prospective prediction setup, not all of the 34 drugs in the study have enough training and test data to build models.

Our work, while offering a proof of the ability to use EHR data for predicting the need for a future drug dose reductions, has several limitations. Dosing changes is part of regular prescription protocol of some drugs and in this case a prediction does not inform patient care. In our study glucocorticoids, such as hydrocortisone and dexamethasone, have guidelines that recommend a dose reduction independently of the drug response. However, for most of the drugs in question (14 out of 23), prescription guidelines recommend dose reduction only in the case of an unwanted reaction. Some drugs we studied (e.g., tacrolimus, warfarin) have guidelines recommending for dose adjustment (up or down). In this case, our system may help by predicting the need for a dose reduction, before first prescription, potentially reducing the necessary time to reach stable dose for these treatments. Another limitation is that we do not account for cases in which adverse drug responses are managed by drug discontinuation or exchanging one drug for another. Accounting for drug discontinuation and switching may improve our performance. We neither consider the co-administration of multiple drugs. Drug prescribed at prediction time, or before, could also be considered as additional feature.

While, we are able to predict dose reductions, the approach fails in predicting dose increases. Our analysis revealed that there are very few conditions mentioned in clinical notes associated with a dose increase for P450-drugs. This lack of observable features for the “absence of a drug response” can be due to lack of reporting in the EHR or to the inability to identify these from EHR. Even for drugs or drug classes associated with good prediction metrics, we observed some false positive and false negative. Those cases are hard to interpret because first of the dimensionality of the model, second because the encoding of the data loose the temporal order of features, making impossible, from this encoding to retrieve the history of a patient. It seems however and unsurprisingly that our model mistakes with patients associated with a lighter density of features.

A third limitation is that while the order of a laboratory test was an important contributor to the performance of the classifier, we have not explored the use of test result values. Result values may enable detection of lack of response to drugs and enable predicting dose increases. Finally, we used a hypergeometric testing with the Bonferroni correction, which is highly stringent, whereas the Holm-Bonferroni¹⁶ is more sensitive.

While our results are promising, more must yet be done to improve the approach. Next generation learning algorithms such as recurrent neural networks may provide substantial performance gains. Considering additional confounding factors may also increase the global performance. Manual chart review or linkage to a biobank with genotype data for the identification of a set of patients with known drug sensitivity could provide further clinical validation. One challenge is also to combine our data-driven approach, with approaches based on broad mechanistic understanding and knowledge of the human physiology as evoked in^{17,18}.

In conclusion, we demonstrate that the phenotypic history of a patient available in the EHR, may be used in machine learning algorithms to identify patients who will likely need a drug dose reduction. Future efforts can focus on a combined strategy for guiding personalized drug prescription, including genetic testing and prior phenotypic data. Several projects such as the Vanderbilt PREDICT, the Mayo Clinic RIGHT programs successfully demonstrate how the genetic determinants in drug response variability may be beneficially used to individualize drug prescriptions and reduce ADR^{19–21}. However, for many situations, the genetic determinants are useful, but not sufficient and may benefit from additional, phenotypic features to guide drug prescription. In this particular case, machine learning approaches such as the one presented in this paper would complement pharmacogenetics testing²². The opportunity is underscored by the fact that roughly 10,620,000 individuals per year in the US receive a new prescription for one of the 34 drugs we built models for (estimated using 2007–2014 Truven MarketScan Commercial Claims and Encounter database). The ability to prospectively identify a large fraction of these individuals (roughly 1,949,000 per year for the 10 drugs for which the dose reduction prediction works well) is highly significant.

Methods

An overview of our approach is presented in Fig. 2. The following section details each step and describes the drugs we consider in this study. The work was done after IRB approval (#24883) at Stanford University, with informed consent for study participation.

Drugs analyzed. We considered in this study drugs whose metabolism is impacted by enzymes of the family of P450 cytochrome, known to interact with many drugs and xenobiotics. We refer to those as *P450-drugs*. Flockhart proposed a list of P450-drugs²³ that is manually reviewed by experts of P450 enzymes and provides elements of evidences in the form of PubMed references. Because there may be too few patients corresponding to a single P450-drug in our data, we also perform the analysis for ‘sets of drugs’. Following the list of Flockhart and the ATC (Anatomical Therapeutic Chemical) classification of drugs²⁴, we identified 25 sets of P450-drugs grouped accordingly to three distinct criteria. To minimize bias related to frequently prescribed drugs we excluded the drugs that were prescribed more than 55,000 times. Of the 205 drugs grouped into 25 drug sets, we also excluded those associated with too few training data, using an arbitrary threshold of 300 drug intervals (≥ 150 dose reductions/increases and ≥ 150 dose continuations). Only 34 drugs across 23 drug sets had enough. Drug sets and their size are summarized in Supplementary Fig. S3, and made available in JSON format in Supplementary file S4.

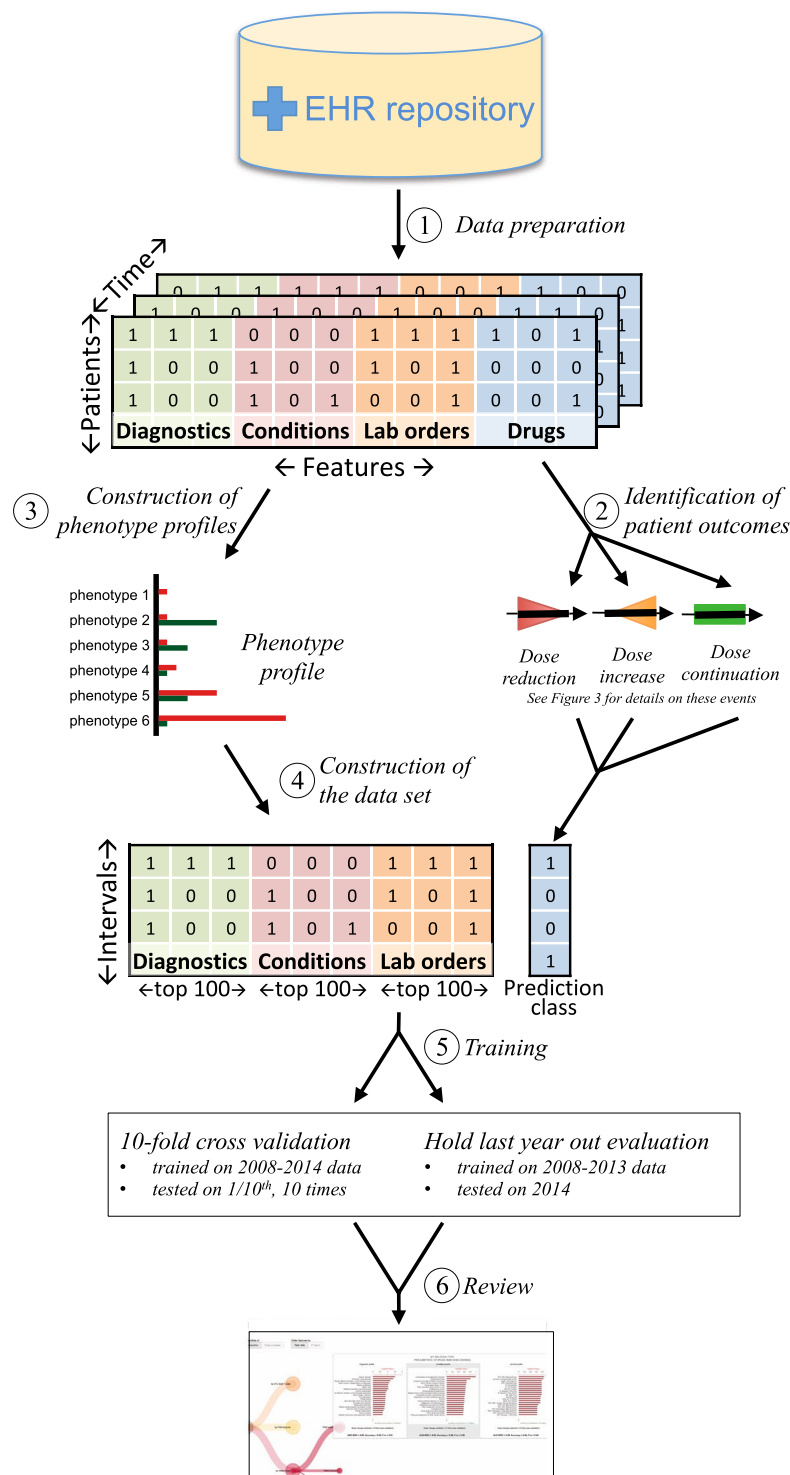


Figure 2. Overview of our approach of predicting the need for a reduced drug dose, at first prescription. (1) Clinical conditions mentioned in clinical notes are identified using terms from medical ontologies; other phenotypic features encoded in Electronic Health Records are directly extracted. (2) Drug dose changes and continuations are detected. (3) Characteristics associated with dose changes are identified to construct phenotype profiles. (4) Phenotype profiles are used to filter the features and build a reduced data set. (5) The resulting matrix is then used to train two binary random forest models: one for predicting dose reductions and one for dose increases. (6) We used two evaluation setups and review by experts to determine the performance of the models.

Dose change and dose continuation intervals. We use EHRs from the STRIDE clinical data warehouse²⁵. It comprises of 1,250,825 patients who visited Stanford Hospital and Clinics between 2008 and 2014, constituting 49,086,060 visits, 27,049,309 clinical notes, 19,435,069 drug orders (including 2,891,470 for P450-drugs)

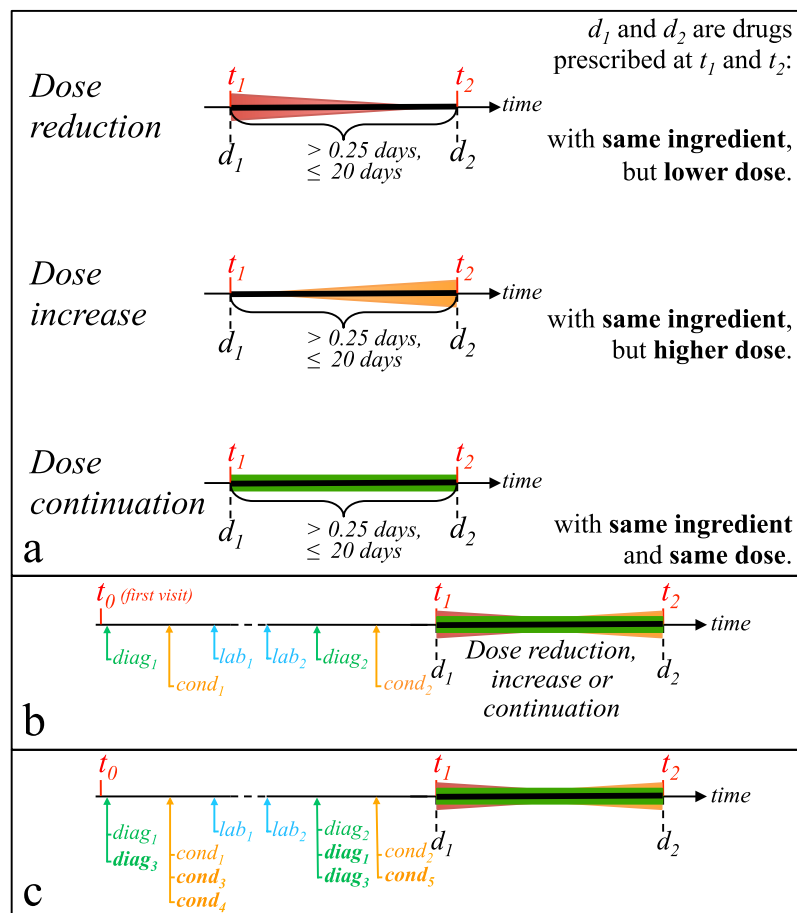


Figure 3. Definition of the dose change intervals and construction of features. The top panel (a) Shows three kinds of intervals. Each interval is delimited by two drug prescriptions d_1 and d_2 , prescribed at t_1 and t_2 . d_1 and d_2 have a same ingredient. No other drug d_i with the same ingredient is prescribed between t_1 and t_2 . The bottom panels show the construction of phenotypic features and their expansion using ontology hierarchies. **(b)** The three kinds of features are: diagnostic codes ($diag_i$), conditions mentioned in clinical notes (p_j) and laboratory test orders (lab_k). Diagnostic codes and test orders are available in EHRs, whereas condition mentions result from automated annotation of clinical notes. **(c)** The diagnostic codes and conditions found in clinical notes are generalized according to ICD-9-CM and SNOMED-CT, respectively. For example, if ICD-9-CM states that $diag_3$ is more general than $diag_1$, then $diag_3$ is also associated with the interval. Generalization is not done for lab tests because too few of them are mapped to an ontology. All features are constructed from EHR data before the first prescription (d_1) of intervals.

and 165,141,675 laboratory test orders. We defined “dose change intervals” as follows: *Dose reductions* are temporal intervals during which a patient was twice prescribed the same drug ingredient with a decreased dose on the second prescription, within 20 days, using the same route of administration and reported with the same unit; *Dose increases* are temporal intervals in which a patient was twice prescribed the same drug ingredient with an increased dose on the second prescription, within 20 days, using the same route of administration and reported with the same unit; Finally, *dose continuations* are intervals between two prescriptions in which the dose, the route and the unit are unchanged. The arbitrary chosen length of 20 days for intervals is supported in our study of P450-drugs by the relatively short length of intervals we observed in EHR (3.64 days, standard deviation = 4.41).

Dose reductions are either a decrease in the quantity of drug prescribed or a decrease of the frequency of prescription. Similarly, dose increases are either an increase of quantity or a increase of frequency. Figure 3 summarizes the three kinds of intervals. We eliminated outlier intervals by excluding the 10% intervals either too short (<6 hours) or too long. In addition, we consider as ‘dose continuations’, only intervals of patients who never experienced a dose change (up or down). However, ‘dose changes’ we consider might precede or follow some dose continuation intervals. We identified 50,704 dose reductions, 60,719 increases and 176,140 continuations in the prescriptions of P450-drugs in STRIDE.

Features observed before dose changes and continuations. We considered three distinct types of features, each observed before the first prescription of the interval (i.e., d_1 in Fig. 3): diagnostic codes, conditions mentioned in clinical notes and laboratory orders.

		Dose reduction	Dose increase	Dose continuation	Total
Intervals		50,704	60,719	176,140	287,563
Patients		22,571	25,381	56,902	69,308
Diagnostic codes	before expansion	1,434,606	1,623,309	4,088,965	7,146,880
	after expansion	3,687,022	4,193,204	10,811,286	18,691,512
Conditions from clinical note	before expansion	441,746	505,235	1,173,403	2,120,384
	after expansion	4,110,928	4,728,006	11,487,221	20,326,155
Lab test orders		6,773,097	7,906,040	17,896,716	32,575,853

Table 3. Numbers of intervals of each type and number of their associated phenotypic features. Diagnostic codes and condition mentions found in clinical notes are expanded using ICD-9-CM and SNOMED CT, respectively.

Diagnostics codes are encoded with the ICD-9-CM (International Classification of Disease, Ninth Revision, Clinical Modification) in EHRs²⁶. ICD-9-CM codes are associated with each visit of a patient at the hospital, documenting the main reasons for the admission, and the main events that occurred during the stay of the patient, reported at discharge.

Conditions mentioned in clinical notes are disease and symptom terms mentioned in clinical narrative notes. They are obtained by processing the clinical notes, using a pipeline described in⁸. This pipeline ignores the mentions of concepts that are negated and those that are mentioned in the patient family history. Its performance for event identification, reported in⁸, are of 74% sensitivity and 96% specificity, but accuracy varies by condition. In this work, we only keep annotations made with concepts that are part of the SNOMED CT²⁷. For the selection of SNOMED CT conditions only, we manually defined a subset of UMLS semantic types²⁸ to consider from: ‘Disease or Syndrome’, ‘Mental or Behavioral Dysfunction’, ‘Cell or Molecular Dysfunction’, ‘Event’, ‘Sign or Symptom’, ‘Anatomical Abnormality’, ‘Neoplastic Process’.

Laboratory test orders are structured data that indicate that a laboratory test, such as *O₂ saturation*, has been ordered for a patient. Unfortunately, our view of the EHR data contains sparse lab results, whereas all orders are listed. While lab orders are not exactly phenotypic observations, they can be considered proxies for a suspected underlying condition. Only 2 in 10 lab orders were encoded with a terminology in our dataset, therefore we used the laboratory order codes directly rather than mappings to a reference terminology.

We used the hierarchies of ontologies to generate an expanded feature set, as illustrated in Fig. 3. If a dose reduction interval is associated with the SNOMED CT concept *Stomatitis*, then it will automatically be associated with parent terms *Inflammatory disorder of digestive tract* and *Disorder of digestive tract* as defined in the SNOMED CT hierarchy. This expansion, on the parent-child hierarchy of the ontology, aims at capturing more information and improving classification of dose changes. Diagnostic codes and conditions are expanded using ICD-9-CM and SNOMED CT respectively, whereas lab orders were not associated with any hierarchy, thus not expanded. Table 3 reports the number of phenotypic features prior and following ontology expansion.

Constructing Phenotype Profiles. We adopt the approach proposed by Lependu *et al.*²⁹ that applied enrichment analyses³⁰ to disease and phenotype studies. Gene expression profiles are commonly computed in transcriptomics to study how gene expression varies depending on conditions such as diseases or treatments. A “profile” in this case is a set of genes differentially (i.e., over- or under-) expressed in one condition compared to another, such as the presence vs. the absence of a disease. Accordingly, gene expression measurements may be replaced by diagnostic codes, condition mentions, or lab test orders in patient EHRs, and the enrichment analysis highlights codes, conditions, or lab tests that are over-represented in a group of patient compared to another.

In this study, we construct phenotype profiles from patients prescribed specific drugs by comparing those who experienced a drug dose reduction with those who had a dose continuation; and comparing patients who experienced a drug dose increase with those who had a dose continuation.

The over-representation of a specific *feature*, is quantified by its *p-value*, computed using the hypergeometric test. A *p-value* quantifies how likely is it that a feature is associated with a dose reduction (or increase) by random chance. We also compute *Risk Ratio* (RR) and *Information Content* (IC) for each feature. The RR measures the strength of association of a feature with the with dose reduction (or increase) of a drug. The IC is used to exclude features that may be either too common or too uncommon in our set of EHRs. Details on the computation of *p-value*, RR and IC are provided in supplementary methods.

Phenotype profiles are initially composed of features with a *p-value* < 0.05 (hypergeometric test). Then, first we remove features with a $0.5 < RR < 2$. Second, features with an IC either in the first or the fourth quartile are excluded. For condition mentions in clinical notes, this step consists of keeping only conditions with $4.25 < IC < 12.75$. Third, only statistically significant associations, after correction for multiple testing are kept ($p < 0.05$, hypergeometric test, Bonferroni correction). Fourth, we applied the *elim* method described by Alexa *et al.*³¹ to deal with features added due to the ontology expansion. One issue with this process is that general concepts from the ontology tend to dominate the set of results. The *elim* method avoids this drawback by eliminating every feature that is more general than another, but is associated with the outcome (dose reduction, increase or continuation) with an equal or higher *p-value*. Table 4 reports numbers of features remaining after each filtering step. The order of the four steps has been chosen for processing efficiency, but could be executed in any order. The resulting phenotype profiles are used to construct train and test sets, which in turn are used to

	Dose reductions			Dose increases		
	diagnostics	conditions	labs	diagnostics	conditions	labs
Total number of features	213,737	92,255	106,041	67,722	23,539	24,231
after RR filtering	104,414	42,225	50,572	41,796	15,122	14,961
after IC filtering	88,973	13,552	34,277	29,875	5,353	5,974
after p-value correction	31,218	3,553	24,442	27,071	5,353	5,974
after elim method	18,256	1,947	—	11,838	1,932	—

Table 4. Reduction of the size of phenotype profiles. This table reports the number of features of each type that comprise phenotype profiles at various steps of filtering. RR and IC filtering steps are based on Risk Ratio (RR) and Information Content (IC). The *elim* method helps filtering results of the ontology expansion process³¹. This method cannot be applied to laboratory test orders since those are not encoded with ontology codes.

learn and evaluate our predictive model. The numbers of publications that come along phenotype profiles are obtained using E-utilities, an API provided by the National Center for Biotechnology Information³². The query for publication numbers is built as the conjunction of two labels: the label associated with the drug (or drug class, see Supplement Material S4 for labels) and the preferred label associated with the feature, following ICD-9 for diagnostics, SNOMED CT for conditions or STRIDE for labs.

Training and evaluation of the predictive model. We trained two binary classifiers using the Random Forest algorithm³³ to predict dose reduction or dose increase as compared to dose continuation. First, phenotype profiles are pruned by selecting features as described before, prior to training. For each classifier, and for each drug and drug set considered, we filtered out two sets of phenotype profiles: a set of profiles with 300 features, composed of the top 100 features of each type (diagnostic code, conditions, lab orders) and a set with profiles composed of all the features that meet statistical significance.

The training sets are balanced, i.e., they include the same number of dose reduction (or increase) intervals as continuations. Features present in the history of patients were encoded with 1, whereas absent or missing features were encoded with 0. Drugs and drug sets with less than 300 instances or with an empty phenotype profile (i.e., no feature was found with sufficient p-value, RR and IC) were excluded from the evaluation; leaving 34 drugs and 23 drug sets to build classifiers for. Every classifier is evaluated with a 10-fold cross-validation and hold last year out validation. In each evaluation setting, the feature selection was achieved on the training data only, meaning that a distinct phenotype profile was computed for each fold of the 10-fold cross-validation. In the hold last year out setting, 2014 data that are left out comprises respectively 17.5%, 18.0% and 18.9% of the dose reduction, increase and continuation intervals of the 2008–2014 data set. Such hold out evaluation illustrates the ability of our model to perform prediction on totally unseen data and also provides the most conservative estimate of performance in the face of data non-stationarity³⁴. Computation of the Random Forest is achieved with the Weka 3.8 toolbox³⁵, with 100 estimators, an unlimited depth of trees, and $\text{int}(\log_2(|\text{features}|) + 1)$ as the number of features considered at each split.

Assessing the association between dose decrease and inappropriate drug response. We manually reviewed prescription drug labels from the DailyMed website (<https://dailymed.nlm.nih.gov>) and when available additional prescription guidelines. From this review we established if it was recommended adjusting the dose of the drug only in the case of inappropriate responses, or if a dose change may be recommended in a normal prescription settings of the drug. Results of the review and references used are provided in Supplement Material S1.

In addition to this review, we counted from clinical notes and for each drug, the number of mentions of conditions classified as Adverse Drug Responses (ADR) according to³⁶, during intervals of time of dose reductions vs. dose continuations, as defined in Fig. 3. We used these counts to compute the *Risk Ratio*, which provides an estimate of the disproportion for ADR to occur during dose reduction intervals in contrast with continuation intervals. Results from this evaluation are reported in Supplementary result S1.

Interpretation of predictions. For interpreting the relationships between features in phenotype profiles and drug sensitivity patterns observed, we (MW, MC, and NJ) reviewed the results and investigated the biomedical literature. To facilitate our interpretation, we developed a web application with the JavaScript library D3JS (<https://d3js.org>), available at <http://snowflake.loria.fr/p450/>.

Ethical approval and informed consent. The work was done with IRB approval (#24883) at Stanford University.

Data Availability

The research uses de-identified EHR data from the Stanford University School of Medicine's clinical data warehouse. Given the regulations governing the access of patient data, the data can not be deposited in a public repository. All other data mentioned in this manuscript are available online as supplementary information. Programmatic code used in this work is available at <https://github.com/coulet/PredictDoseReduction>.

References

1. Jameson, J. L. & Longo, D. L. Precision Medicine — Personalized, Problematic, and Promising. *N. Engl. J. Med.* **372**, 2229–34 (2015).
2. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. National Action Plan for Adverse Drug Event Prevention. Washington, DC (2014).
3. Bourgeois, F. T., Shannon, M. W., Valim, C. & Mandl, K. D. Adverse drug events in the outpatient setting: an 11-year national analysis. *Pharmacoepidemiol Drug Saf.* **19**, 901–10 (2010).
4. Stewart, W. F., Shah, N. R., Selna, M. J., Paulus, R. A. & Walker, J. M. Bridging the inferential gap: the electronic health record and clinical evidence. *Health Aff.* **26**, w181–w191 (2007).
5. Jensen, P. B., Jensen, L. J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* **13**, 395–405 (2012).
6. Miotto, R., Li, L., Kidd, B. A. & Dudley, J. T. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports* **6**, 26094 EP (2016).
7. Tatonetti, N. P., Ye, P. P., Daneshjou, R. & Altman, R. B. Data-driven prediction of drug effect and interactions. *Sci. Transl. Med.* **4**, 125ra131 (2012).
8. LePendou, P. *et al.* Pharmacovigilance using clinical notes. *Clin. Pharmacol. Ther.* **93**, 547–555 (2013).
9. Neuraz, A. *et al.* Phenome-wide association studies on a quantitative trait: application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. *PLoS Comput. Biol.* **9**, e1003405 (2013).
10. Shah, N. H. Mining the ultimate phenome repository. *Nat Biotechnol.* **31**, 1095–1097 (2013).
11. Ramirez, A. H. *et al.* Predicting warfarin dosage in European-Americans and African-Americans using DNA samples linked to an electronic health record. *Pharmacogenomics* **13**(4), 407–418 (2012).
12. Delaney, J. T. *et al.* Predicting clopidogrel response using DNA samples linked to an electronic health record. *Clin. Pharmacol. Ther.* **91**(2), 257–263 (2012).
13. Birdwell, K. A. *et al.* The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients. *Pharmacogenet. Genomics* **22**(1), 32–42 (2012).
14. Samwald, M. *et al.* Incidence of Exposure of Patients in the United States to Multiple Drugs for Which Pharmacogenomic Guidelines Are Available. *PLoS ONE* **11**, e0164972 (2016).
15. Danielson, P. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.* **3**, 561–597 (2002).
16. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics.* **6**, 65–70 (1979).
17. Tsamandouras, N., Rostami-Hodjegan, A. & Aarons, L. Combining the “bottom up” and “top down” approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. *British Journal of Clinical Pharmacology* **79**(1), 48–55 (2015).
18. Darwich, A. S. *et al.* Why Has Model-Informed Precision Dosing Not Yet Become Common Clinical Reality? Lessons From the Past and a Roadmap for the Future. *Clin. Pharmacol. Ther.* **101**, 646–656 (2017).
19. Roden, D.M. *et al.* Benefit of Preemptive Pharmacogenetic Information on Clinical Outcome. *Clin Pharmacol Ther.* **cpt.1035** (2018).
20. Volpi, S. *et al.* Research Directions in the Clinical Implementation of Pharmacogenomics: An Overview of US Programs and Projects. *Clin Pharmacol Ther.* **cpt.1048** (2018).
21. Denny, J.C., Van Driest, S.L., Wei W.Q. & Roden, D.M. The Influence of Big (Clinical) Data and Genomics on Precision Medicine and Drug Development. *Clin Pharmacol Ther.* **cpt.951** (2018).
22. Ioannidis, J. P. A. To replicate or not to replicate: The case of pharmacogenetic studies. *Circulation: Cardiovascular Genetics* **6**, 413–418 (2013).
23. Flockhart, D. Drug Interactions: Cytochrome P450 Drug Interaction Table. Indiana University School of Medicine at, <http://medicine.iupui.edu/clinpharm/ddis/> [visited June 6th, 2018] (2007).
24. World Health Organisation Collaborating Center for Drug Statistics Methodology, ATC structure and principles at, https://www.whocc.no/atc/structure_and_principles/ [visited June 6th, 2018] (2018).
25. Lowe, H., Ferris, T., Hernandez, P. & Weber, S. STRIDE—An integrated standards-based translational research informatics platform. *In AMIA Annu. Symp. Proc.* **2009**, 391–395 (2009).
26. Centers for Disease Control and Prevention, International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) at, <https://www.cdc.gov/nchs/icd/icd9cm.htm> [visited June 6th, 2018] (2015).
27. SNOMED International, SNOMED CT web page. at, <https://www.snomed.org/snomed-ct> [visited June 6th, 2018].
28. McCray, A. T. *et al.* An Upper Level Ontology for the Biomedical Domain. *Comp. Funct. Genom.* **4**, 80–4 (2003).
29. LePendou, P., Musen, M. A. & Shah, N. H. Enabling enrichment analysis with the Human Disease Ontology. *J. Biomed. Inform.* **44**, S31–38 (2011).
30. Khatri, P. & Drăghici, S. Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* **21**, 3587–3595 (2005).
31. Alexia, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–7 (2006).
32. Sayers, E. A. General introduction to the E-utilities. In: *Entrez Programming Utilities Help*. National Center for Biotechnology Information, at: <https://www.ncbi.nlm.nih.gov/books/NBK25497/> [visited June 6th, 2018] (2010).
33. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32.
34. Jung, K. & Shah, N. H. Implications of non-stationarity on predictive modeling using EHRs. *J. Biomed. Inform.* **58**, 168–74 (2015).
35. Frank, E., Hall, M. A. & Witten, I. H. The WEKA Workbench. Online Appendix for *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Fourth Edition (2016).
36. Staussberg, J. & Hasford, J. Drug-related admissions and hospital-acquired adverse drug events in Germany: a longitudinal analysis from 2003 to 2007 of ICD-10-coded routine data. *BMC Health Services Research* **11**, 134 (2011).

Acknowledgements

This work is supported by: The Snowflake/Snowball Associate Team funded by Inria; The PractiKPharma project (ANR-15-CE23-0028) funded by The French Research Agency (ANR); The Widen Horizons program of the IDEX “Lorraine Université d’Excellence” (15-IDEX-0004); The NIGMS grant R01 GM101430 and NLM grant R01 LM011369. We thank Juan H. Banda, Alison Callahan, Rohit Vahisht and Rainer Winnenbourg for inputs in data processing, and IC calculation.

Author Contributions

A.C., M.D., N.H.S. conceived the study. A.C., N.J., N.H.S. designed the analysis methods. A.C. implemented the analysis, and generated the results. A.C., M.D., N.H.S. drafted the manuscript. M.B.C., M.W. and N.J. reviewed the phenotype profiles that led to high predictive accuracy. All authors have read, and approve of the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33980-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018