PrOCTOR: A Data-Driven Model for Predicting Drug Toxicity and Its Ethical Implications in Clinical Trials

Licza Lobo

October 25, 2024

A prescribed drug's journey begins with a rigorous development process regulated by the U.S. Food and Drug Administration (FDA) ("What Are Clinical Trials and Studies?"). This process requires both a substantial amount of time, about 10 to 15 years, and high costs, around \$1 billion for each successful drug ("What Are Clinical Trials and Studies?"). Clinical trials are a key component of this process, allowing for human testing to develop new treatments, improve upon existing medications, prevent health issues, and improve the quality of life for the sick population ("What Are Clinical Trials and Studies?"). The FDA requires that drugs first pass a preliminary discovery phase, followed by phases 1, 2, and 3 trials to assess their safety and effectiveness ("What Are Clinical Trials and Studies?"). However, approximately 90% of all drugs fail clinical trials across all phases ("What Are Clinical Trials and Studies?"). 30% of these failures are attributed to unmanageable toxicity levels that lead to serious adverse effects in humans (Sun). Failure rates have continued to rise over the past years, meaning an increase in risks to clinical trial participants, the delay of medical advancements, and the 'waste' of research funding ("What Are Clinical Trials and Studies?" , Sun). This has prompted many to seek improvements in the clinical trial process (Gayvert et al.). This paper will analyze the methodologies and results presented in "A Data-Driven Approach to Predicting Successes and Failures of Clinical Trials," where the authors develop a data-driven model to predict the likelihood of drug toxicity in clinical trials, alongside the ethical implications this raises (Gayvert et al.).

To begin with the paper's methodology, the authors examined current screening measures for identifying drugs likely to fail clinical trials. The primary methods, known as rule-based methods, for eliminating toxic drug candidates in the early stages rely solely on "drug likeness," a measure of how molecularly similar drug candidates are to FDA approved drugs. These methods include Lipinski's Rules of Five, Veber's Method, Ghose's Rule, and the quantitative estimate of drug likeness (QED). To test their accuracy, the researchers compared 108 drugs that failed clinical trials due to toxicity, from ClinicalTrials.gov, against 1,013 FDA-approved drugs from the DrugBank database. The findings revealed that both classes of drugs passed Lipinski's, Veber's, and Ghose's rules at comparable rates, with Veber's method being the most conservative, predicting failure for 75.2% of approved drugs and 92% of failed drugs. Additionally, a Kolmogorov-Smirnov test indicated the QED method showed no significant ability to differentiate between the two classes of drugs. These results

suggest that a sole focus on chemical properties does not accurately classify potentially toxic drugs. Therefore the researchers proposed a more comprehensive model that incorporates additional features related to drug performance, specifically target-based properties.

Using the DrugBank database, the study generated a set of target-based properties, including gene tissue expression, network connectivity, and loss-of-function frequency. These properties were then integrated with molecular and drug likeness features validated by currently used methods resulting in 48 features (10 molecular, 34 target-based, and 4 related to drug likeness). Kolmogorov-Smirnov tests determined that both chemical and target-based properties could individually discriminate between approved and failed drugs, producing weak but statistically significant results. A correlation analysis showed that the 28 target gene expression properties were highly correlated, leading the researchers to use Principal Component Analysis to reduce the predictors. This analysis narrowed the target gene expression measures down to the first three principal components. Lastly, the study assessed the Pearson correlation between chemical and target-based feature subsets, finding no significant correlation, r=0.194. This lack of correlation indicates the target-based features are independent of chemical properties, validating the theory that they can add information when predicting drug toxicity. These methods resulted in a final total of 20 predictors, 10 chemical and 10 target-based features, for the development of PrOCTOR, a model that predicts the odds of drug toxicity in clinical trials.

Researchers utilized a random forest model approach when training PrOCTOR, using the 108 failed drugs and 1,013 FDA approved drugs as the training dataset. Sub-sampling methods (30 iterations to create 50 bootstrapped subsets) were implemented to address the imbalance ratio of failed to approved drugs in the training set. This approach minimized the risk of biased samples. Each of the 50 subsets resulted in a decision tree with a random subset of splits, yielding a class label and associated probability. The researchers determined an average probability to derive an odds score reflecting the likelihood of approval versus failure. The PrOCTOR score is then defined as the log2 of this odds score.

10-fold cross validation was conducted on PrOCTOR using a set of 784 FDA approved drugs and associated 100 failed drugs. This produced an AUC of 0.8263, indicating significant predictive power. Researchers then compared the optimal point of the curve and true positive rate (TPR) of 0.7544 to the TPR of the rule-based methods with the same dataset. PrOCTOR's TPR was comparable to that found by Lipinski's method and greater than those of Ghose and Veber. ROC curve comparisons and the Wilcoxon Signed Rank Test demonstrated that PrOCTOR had significantly greater predictive power than the QED method. By comparing PrOCTOR to currently used methods, the research team ensured the model's reliability as a drug evaluation tool.

Further model development and validation involved applying PrOCTOR to non-FDA drugs approved in Europe and Japan. The distribution of PrOCTOR scores for these drugs was similar to that of FDA-approved drugs. Mann-Whitney U tests revealed significant differences in classification between foreign-approved drugs and failed drugs in the training set, indicating that drugs approved in Europe and Japan were predicted to be significantly safer. Next, the model was applied to 3,236 unseen FDA approved drugs from the Drugbank database. Unpaired Student's t-test revealed a significant difference in serious adverse

events between drugs predicted as safe and toxic. Similarly, when applied to a set of 137 drugs labeled as high concern and 65 labeled as no concern, Mann-Whitney tests (p=0.005) indicated that FDA-approved drugs with higher concern had higher odds of being classified as toxic. The extensive application of the model to new datasets demonstrates PrOCTOR's high generalizability and effectiveness in predicting drug toxicity.

To conclude the testing of the developed PrOCTOR model, researchers further analyzed six compounds identified as the most and least toxic by the model, taking advantage of the fact PrOCTOR produces both a class label and continuous score measure. The compounds identified were mostly consistent with rule-based methods and current literature. However, Rosiglitazone, identified as one of the worst compounds by PrOCTOR but approved by rule-based methods, was later removed from the market due to high risks of heart attack. This suggests that PrOCTOR may predict future adverse clinical outcomes, unlike traditional methods. To evaluate PrOCTOR's effectiveness in this context, researchers compared PrOCTOR scores to SIDER side effect scores, expecting to see a negative relationship, where better PrOCTOR score correlates with less severe side effects. Spearman rank correlation coefficients assessed the relationship between severe side effects and predicted toxic drugs and Fisher's exact tests were conducted to compare side effects in the 90th and 10th percentiles of PrOCTOR scores. The results demonstrated that compounds identified as more toxic had worse PrOCTOR scores, reinforcing the model's predictive capabilities regarding safety.

Through this analysis, researchers were successfully able to create a model that combines chemical and target-based properties to improve upon existing predictive methods. By demonstrating its generalizability to other datasets and relationships to adverse clinical outcomes, the researchers have shown PrOCTOR's potential as a robust model to reduce clinical trial failure rates and improve drug development efficiency. However, what ethical implications does this hold for the future of drug development?

Integrating machine learning models into clinical decision-making has significant real-world implications not only for those involved in drug development but also for the populations relying on new drugs to improve their quality of life. Utilizing PrOCTOR predictions to determine which drug candidates to pursue may overshadow the judgment and expertise of trained clinicians. Clinicians consider a variety of factors when evaluating drug candidates, including the specifics of clinical trial design, the drugs themselves, the patients impacted, and the conditions targeted. As medicine continues to shift towards more personalized approaches, models like PrOCTOR may struggle to capture these complexities (Husnain et al.). Furthermore, the reliance on data-driven approaches can lead to an overdependence on model predictions, which can potentially limit the development of future beneficial treatments. This raises the question of whether such impactful decisions should rest in the hands of data-driven models. Are clinicians justified in their use of the model to decide on drug candidates?

When attempting to answer this question it is important to consider the actual patients who benefit from or take these drugs. This raises ethical concerns related to data privacy, patient autonomy, and potential delays in access to treatments. Models that predict success in clinical trials rely on large datasets, including patient records, clinical trial histories, and

genomic information (Husnain et al.). The use of such data must align with data protection laws and require informed patient consent. However, for many patients participating in trials, especially those facing severe illness or end-of-life situations, obtaining consent may not be feasible. This raises the question: does the use of their data violate consent laws? Furthermore, a shift from a patient-centered approach to data-driven decision-making can diminish the voice of patients. Many individuals lack an understanding of the algorithms that inform medical decisions. This reduction in transparency can decrease trust in science and make patients feel disconnected from the processes that affect their health (Husnain et al.). Such models also fail to consider patients' individual needs, perspectives, and preferences. Therefore, the integration of machine learning in the pharmaceutical industry may lead to a reduced emphasis on patient perspectives in favor of model predictions, raising ethical concerns.

These issues can be further examined in terms of their differential impact on various patient populations. When analyzing drug properties to determine success, certain drugs may be overlooked despite their potential to benefit specific patients, particularly those suffering from rare or extremely severe conditions. These conditions often have a lower threshold for negative effects, making model predictions critical. However, this can lead to a reduction in viable treatment options for such patients, disproportionately affecting those with severe illnesses (Husnain et al.). Is it fair for patients suffering from underrepresented diseases to have fewer treatment options due to model predictions? This question highlights the need for careful consideration of equity in drug development and the ethical implications of relying on data-driven approaches that may not adequately address the needs of all patient populations.

The development of the PrOCTOR model by Gayvert and contributing researchers highlights the significant potential of predictive tools to enhance the safety and efficiency of drug development. These models can improve clinical outcomes by reducing the risk of failed drug studies and providing clinicians with valuable insights to monitor long-term adverse effects. However, as these tools are integrated into the medical field, it is crucial to ensure they do not overshadow patient and clinical judgment. Medical decisions should remain a collaborative dialogue between providers and patients, creating personalized approaches to care. The introduction of predictive models may shift medical decision-making into the hands of data, overpowering trained clinicians and neglecting patient preferences. Thus, this normative concern is pressing, highlighting the need to balance the benefits of these models with their ethical implications to ensure that patient-centered care remains a priority.

References

Gayvert, Kaitlyn M., et al. "A Data-Driven Approach to Predicting Successes and Failures of Clinical Trials." Cell Chemical Biology, vol. 23, no. 10, Oct. 2016, pp. 1294–1301, https://doi.org/10.1016/j.chembiol.2016.07.023.

Husnain, Ali, et al. "View of Revolutionizing Pharmaceutical Research: Harnessing Machine Learning for a Paradigm Shift in Drug Discovery." International Journal of Multidisciplinary Sciences and Arts, vol. 2, no. 2, Dec. 2023, jurnal.itscience.org/index.php/ijmdsa/article/view/2897/2219. Accessed 23 Oct. 2024.

Sun, Duxin. "90% of Drugs Fail Clinical Trials." ASBMB, 12 Mar. 2022, www.asbmb.org/asbmb-today/opinions/031222/90-of-drugs-fail-clinical-trials. Accessed 25 Oct. 2024.

"What Are Clinical Trials and Studies?" National Institute on Aging, 22 Mar. 2023, www.nia.nih.gov/health/clinical-trials-and-studies/what-are-clinical-trials-and-studies.