



Rise of the Machines? Predicting Brivaracetam Response Using Machine Learning

Epilepsy Currents
2022, Vol. 22(2) 111–113
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15357597211049052
journals.sagepub.com/home/epi
 SAGE

de Jong et al. *Brain*. 2021;144:1738–1750.

Accurate and individualized prediction of response to therapies is central to precision medicine. However, because of the generally complex and multifaceted nature of clinical drug response, realizing this vision is highly challenging, requiring integrating different data types from the same individual into one prediction model. We used the anti-epileptic drug brivaracetam as a case study and combine a hybrid data/knowledge-driven feature extraction with machine learning to systematically integrate clinical and genetic data from a clinical discovery dataset ($n = 235$ patients). We constructed a model that successfully predicts clinical drug response [area under the curve (AUC) = .76] and show that even with limited sample size, integrating high-dimensional genetics data with clinical data can inform drug response prediction. After further validation on data collected from an independently conducted clinical study (AUC = .75), we extensively explore our model to gain insights into the determinants of drug response and identify various clinical and genetic characteristics predisposing to poor response. Finally, we assess the potential impact of our model on clinical trial design and demonstrate that, by enriching for probable responders, significant reductions in clinical study sizes may be achieved. To our knowledge, our model represents the first retrospectively validated machine learning model linking drug mechanism of action and the genetic, clinical and demographic background in epilepsy patients to clinical drug response. Hence, it provides a blueprint for how machine learning-based multimodal data integration can act as a driver in achieving the goals of precision medicine in fields such as neurology.

Commentary

If we knew who would respond to which treatments ahead of time, our jobs would be considerably easier. Sadly, we do not. Furthermore, ‘gold standard’ randomized controlled trials assess only average treatment effects within a groomed population, oftentimes unpowered for one-at-a-time subgroup analyses.

Enter machine learning classification models, which have been used widely in epilepsy¹ and can break free of restrictive regression assumptions. The concept of individualized prediction is alluring. Even an experienced clinician’s gestalt is not perfect,^{2,3} and ‘big data’ offers the opportunity to more precisely forecast treatment responses better than averages alone and better than human predictions alone (here is one recent example using gradient-boosted decision trees⁴).

de Jong et al⁵ recently asked, can we predict who responds to brivaracetam? First, they acquired data⁶ randomizing 760 adults with ongoing focal seizures 1:1:1 to placebo or brivaracetam (100 mg/day or 200 mg/day). They defined ‘responders’ as >50% reduction in seizure frequency at 12 weeks compared to baseline (22% placebo vs ~38% brivaracetam groups). Second, they developed 4 ‘modalities’ (aka groups of predictors): 1) gene set-wise mutational load scores, 2) polygenic risk scores, 3) SV2A variants and 4) 106 clinical features. They started with just about every known gene related to epilepsy or brivaracetam (~20 million single nucleotide variants!), whittled to a mere 14 000.

Third, for the 235/497 brivaracetam-treated patients who gave blood samples, they plugged these predictors into 5 machine learning models to see which model would come out on top. Models were judged based on the ‘area under the curve’ (AUC: ‘the probability that a random subject who experienced the outcome had a higher predicted probability than one who did not), where .5 is chance and 1 is perfect. Some models relied upon linear assumptions (eg 1. discriminant analysis which reduces a gargantuan number of variables into a smaller number of dimensions separating responders from non-responders and 2. elastic net, like a typical multivariable regression except also nudging coefficients towards 0 to reduce overfitting in-sample noise). Others were nonlinear (eg 1. gradient-boosted decision trees, like a plinko board iteratively shrinking residuals and 2. neural networks).

They found the following:

- 1) Decision tree models (AUC .72-.76) outperformed the others (AUC .63-.67). Considering any of the 3 genetic modalities alone resulted in poor classification (AUC .55-.58), considering clinical information alone resulted in acceptable classification (AUC .71) and the ‘integrated model’ combining all 4 modalities produced the strongest classification (AUC .76). Basically, ability to predict short-term ‘>50% reduction’ using all available clinical plus genetic data was reasonable, though not quite strong.





- 2) The single most important variable predicting outcomes was prior levetiracetam treatment (worse), and the next 5 most important predictors were also clinical (eg seizure frequency and anxiety), followed by a microtubule binding protein mutation, followed lower down the list by SV2A variants.

So, have artificially intelligent machines hungry for whole-genome sequences risen to finally replace clinical intuition? While the authors and trialists have made heroic gains to complete this computationally intensive work, my answer remains – not yet.

First, despite the hype surrounding machine learning, caution is required. In assumption-free models fed an enormous number of features, overfitting in-sample noise is a real concern. However, de Jong et al did find (truly surprisingly) good external validation (AUC .75) against 47 patients from a second RCT.⁷ Even so, it is an uphill battle getting clinicians to trust complex unfamiliar methods, using either routinely available variables classified in non-routine ways or else non-routine variables, without clearly demonstrating superiority over simpler models. Another crucial point is that, machine learning or not, no model is exonerated from underlying data limitations. For example, it is not clear how meaningful a >50% relative seizure reduction is with only a small absolute difference between groups (median 1.75 vs 1.26 seizures/week at follow-up; seizure-freedom rate of 4% in the 100 mg/day group⁷), the data are based on refractory patients rather than all-comers, the selection process determining which patients contributed blood samples seems unclear, a ‘responder’ rate among the treated does not disentangle placebo effect and enrolling refractory epilepsy implies regression to the mean (the most severe cases tend to ‘respond’ towards average no matter what you do).

Second, individualized treatment prediction incorporating genetics was another source of possible hype. Indeed the ‘integrated model’ encompassing extensive genetics plus clinical information (AUC .76) outperformed a model with clinical information alone (AUC .71). However, despite incredible statistical significance ($P = .0000019$) owing to an enormous feature space, improvements were slim – a boost in discrimination of .05 from genome-wide spelunking is not headline news, especially considering the currently huge expense, ambiguity and delay entailed in obtaining whole-genome sequencing. SV2A variants, the most targeted hypothesis here, also predicted outcomes only trivially better than chance (AUC .58) and substantially worse than clinical information alone (.71). Even thousands of additional SNPs chosen specifically for their relevance to epilepsy resulted in virtually no better than chance discrimination (AUC .55). From this author reconstructing Figure 4B,⁶ considering only treatment group and prior levetiracetam in a mere 2-variable logistic regression produces an AUC .64, better than any presented machine learning genetics-based modality.

Third, even if the AUC had been 1, it would be a mistake to declare ‘mission accomplished’. To be useful (‘all models are wrong, but some are useful’), a model must not only demonstrate external validity but also be disseminated to and usable by

others (the authors do not show their actual decision tree for others to use) and should estimate outcome probabilities comparing treatment versus no treatment (showing only what might happen under ‘treatment’ lacks the key comparison to ‘no treatment’). And even then, it would remain far from given that the right clinical answer aligns with the ‘predicted’ answer. Just because a patient is predicted to have a <50% relative response rate does not rule out benefit, nor does predicting a >50% response rate necessarily rule in treatment. Even if we perfectly predicted this particular outcome, the next step remains entailing sometimes a complex net-benefit analysis balancing absolute (not just relative) benefits and harms (not just benefits).

Estimating individualized treatment responses ahead of time is a central goal for twenty first century precision medicine. I do want to emphasize the usefulness of data-driven predictive models intended to assist the clinician (some existing examples^{8,9}) – we need more, and de Jong et al have conducted important (and hard) work. Next steps could be larger training datasets capturing more variation to improve predictiveness, and future changes in the genetics landscape (discovering more genes, cost reductions) could further change the game. Still, we are currently far from having ‘realized the vision of precision medicine’, and clinicians are still often left with the conventional trial and error approach to antiseizure medications.

By Samuel W. Terman, MD MS

Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. *Epilepsia*. 2019;20:2037-2047.
2. Hoffmann TC, Del Mar C. Clinicians’ expectations of the benefits and harms of treatments, screening, and tests: A systematic review. *JAMA Int Med*. 2020;4229(3):407-419.
3. Gracia CG, Chagin K, Kattan MW, Ji X, Kattan MG, Crotty L, et al. Predicting seizure freedom after epilepsy surgery, a challenge in clinical practice. *Epilepsy Behav [Internet]*. 2019;95:124-130. Available from: DOI: [10.1016/j.yebeh.2019.03.047](https://doi.org/10.1016/j.yebeh.2019.03.047).
4. van Doorn WPTM, Stassen PM, Borggreve HF, Schalkwijk MJ, Stoffers J, Bekers O, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. *PLoS One [Internet]*. 2021;16(1):1-15. Available from: DOI: [10.1371/journal.pone.0245157](https://doi.org/10.1371/journal.pone.0245157).
5. de Jong J, Cutcutache I, Page M, Elmoufti S, Dilley C, Fröhlich H, et al. Towards realizing the vision of precision medicine: AI based prediction of clinical drug response. *Brain*. 2021;144(6):1738-1750.
6. Klein P, Schiemann J, Sperling MR, Whitesides J, Liang W, Stalvey T, et al. A randomized, double-blind, placebo-controlled, multicenter, parallel-group study to evaluate the efficacy and safety of adjunctive

-
- 
- brivaracetam in adult patients with uncontrolled partial-onset seizures. *Epilepsia*. 2015;56(12):1890-1898.
7. Ryvlin P, Werhahn KJ, Blaszczyk B, Johnson ME, Lu S. Adjunctive brivaracetam in adults with uncontrolled focal epilepsy: Results from a double-blind, randomized, placebo-controlled trial. *Epilepsia*. 2014;55(1):47-56.
 8. AED Withdrawal Risk Calculator. Available from: <http://epilepsypredictiontools.info/aedwithdrawal>. <http://epilepsypredictiontools.info/aedwithdrawal>. Accessed September 7, 2021.
 9. Cleveland Clinic Risk Calculator Library. Department of Quantitative Health Sciences. Available from: riskcalc.org Accessed September 7, 2021.