Check for updates

MINI-REVIEW

Machine Learning in Modeling Disease Trajectory and Treatment Outcomes: An Emerging Enabler for Model-Informed Precision Medicine

Nadia Terranova[1,*] and Karthik Venkatakrishnan[2,*]

The increasing breadth and depth of resolution in biological and clinical data, including -omics and real-world data, requires advanced analytical techniques like artificial intelligence (AI) and machine learning (ML) to fully appreciate the impact of multi-dimensional population variability in intrinsic and extrinsic factors on disease progression and treatment outcomes. Integration of advanced data analytics in Quantitative Pharmacology is crucial for drug-disease knowledge management, enabling precise, efficient and inclusive drug development and utilization – an application we refer to as model-informed precision medicine. AI/ML enables characterization of the molecular and clinical sources of heterogeneity in disease trajectory, advancing end point qualification and biomarker discovery, and informing patient enrichment for proof-of-concept studies as well as trial designs for efficient evidence generation incorporating digital twins and virtual control arms. Explainable ML methods are valuable in elucidating predictors of efficacy and safety of pharmacological treatments, thereby informing response monitoring and risk mitigation strategies. In oncology, emerging opportunities exist for development of the next generation of disease models via ML-assisted joint longitudinal modeling of high-dimensional biomarker data such as circulating tumor DNA and radiomics profiles as predictors of survival outcomes. Finally, mining real-world data leveraging ML algorithms enables understanding of the impact of exclusion criteria on clinical outcomes, thereby informing rational design of appropriately inclusive clinical trials through data-driven broadening of eligibility criteria. Herein, we provide an overview of the aforementioned contexts of use of ML in drug-disease modeling based on examples across multiple therapeutic areas including neurology, rare diseases, autoimmune diseases, oncology and immuno-oncology.

Model-informed precision medicine[1] is envisioned as the application of quantitative pharmacology models to generate hypotheses and/ or substantiate evidence regarding the contribution of patient-specific factors (e.g., genetic variation in the molecular target of a drug) to the efficacy and/ or safety of therapeutics. Model-informed precision medicine takes the established discipline of model-informed drug development into the realm of informing not only the development but also the optimal clinical use of precision medicines, ensuring the utilization of the right drug for the right patient, and for all patients. With a steady increase in the dimensionality of biological and clinical data, traditional clinical pharmacology study designs and methods of covariate analysis in pharmacometrics to evaluate intrinsic/ extrinsic factor effects fall short when faced with big "omics" data sets. These include patient-level characterizations of the genome, transcriptome, proteome, metabolome, radiome, microbiome, and exposome. Furthermore, there is growing utilization of real-world data (RWD) from diverse healthcare settings. To address these challenges, artificial intelligence (AI)/machine learning (ML) and related advanced analytical approaches have rapidly permeated the disciplines of Pharmacometrics and Quantitative Systems Pharmacology. These techniques enable integration of multimodal and multidimensional sources of variability in disease trajectory and the pharmacological response to therapeutics.[2] Such integration is vital across the pharmaceutical research and development lifecycle. It helps inform confidence in the biological target beginning early in drug discovery, selection of the right patient population for clinical development, and maximization of benefit versus risk through optimized dosage selection for all patients.[3] Herein, we provide an overview of opportunities for integrating ML in drug-disease modeling for characterizing disease progression, variability in treatment response and enhancing diversity and inclusion in clinical development (**Figure 1**).

CHARACTERIZING HETEROGENEITY IN DISEASE TRAJECTORY

Many chronic diseases (e.g., autoimmune, cardiovascular, neurological, and neurodegenerative diseases) have complex pathophysiology, translating to substantial heterogeneity in disease trajectory, and inter-patient variability in treatment outcomes. In

[1]Quantitative Pharmacology, Ares Trading S.A. (an affiliate of Merck KGaA, Darmstadt, Germany), Lausanne, Switzerland; [2]EMD Serono Research and Development Institute, Inc., Billerica, Massachusetts, USA. *Correspondence: Nadia Terranova (nadia.terranova@emdgroup.com); Karthik Venkatakrishnan (karthik.venkatakrishnan@emdserono.com)

Received October 13, 2023; accepted December 11, 2023. doi:10.1002/cpt.3153

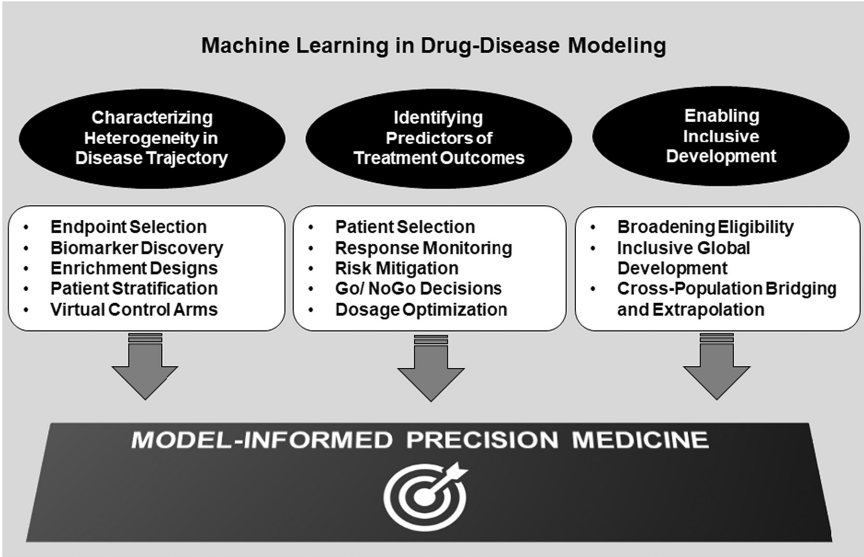
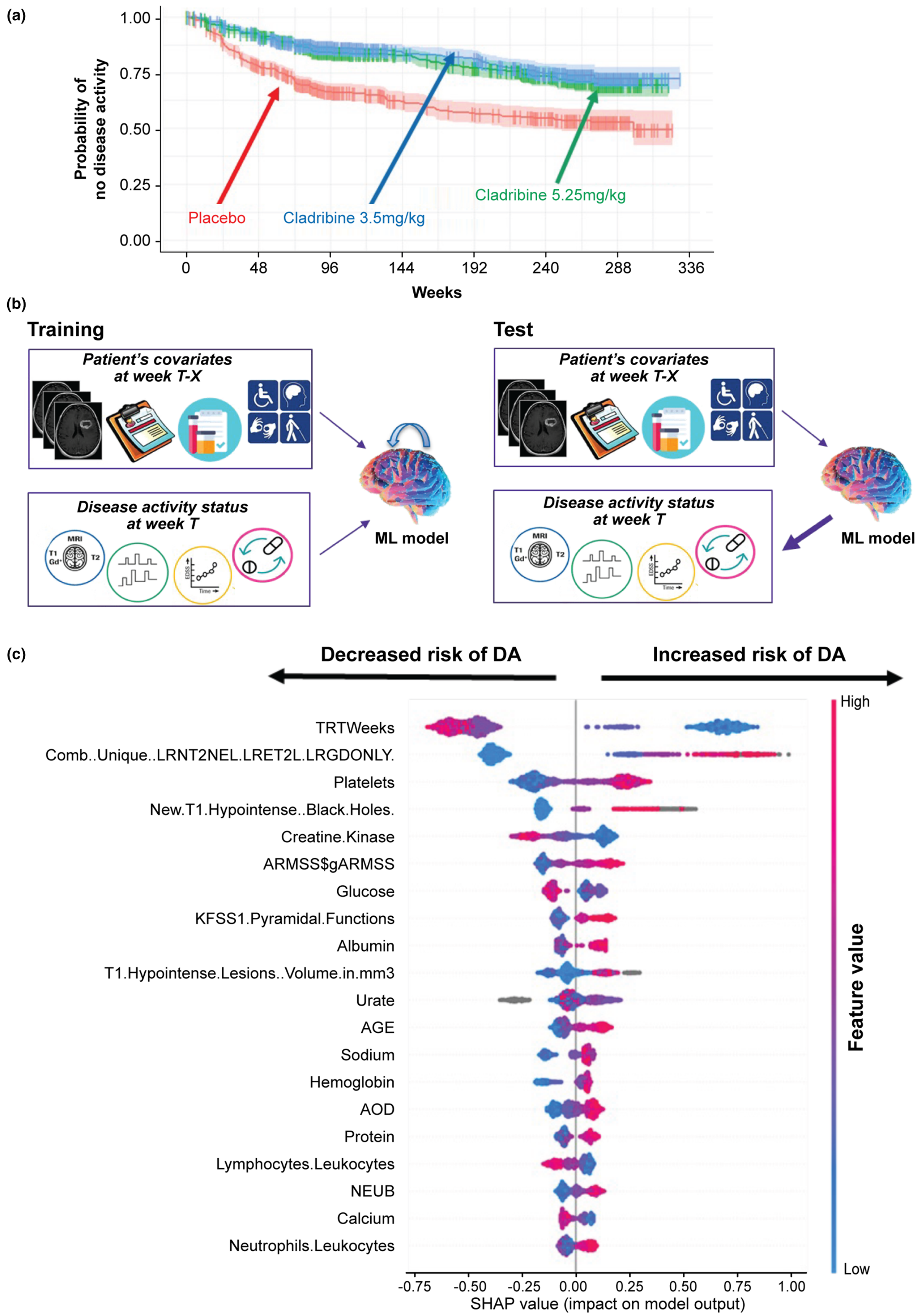


Figure 1 Opportunities for integrating machine learning in drug-disease modeling for model-informed precision medicine.

autoimmune diseases like systemic lupus erythematosus (SLE), the relationships between soluble biomarker profiles (e.g., autoantibodies) and disease pathophysiology or treatment outcomes are complex. These relationships and are often poorly understood to inform patient selection/ stratification hypotheses and relationships to treatment outcomes. Furthermore, understanding of the time course of changes in such biomarkers in relation to the time course of clinical outcomes based on longitudinal patient-level data is often absent. Attrition in drug development in these disease areas due to lack of efficacy in phase II trials despite scientifically appealing therapeutic hypotheses remain a significant challenge. The multidimensionality of longitudinal biomarker data linked to clinical outcomes motivates the application of ML approaches to reveal predictors of disease trajectory and treatment outcomes going beyond traditional cross-sectional analyses that often ignore considerations of the time course of pathophysiological and pharmacological processes, as well as the underlying interactions and relationships among those. In a recent example, a longitudinal clustering method was applied to data from a cohort of 805 patients with SLE to identify predictive profiles for clinical outcomes.[4] The study analyzed data from 805 patients with SLE over 10 years, identifying 4 distinct patient clusters based on autoantibody profiles. These clusters were associated with differences in disease activity, organ involvement, treatment requirements, and mortality risk, which also resulted in differences in survival. Analyses of this nature have the potential to inform intelligent proof-of-concept (POC) trial designs through appropriate patient stratification going beyond traditional phenotypic traits, ultimately contributing to enhancing probability of technical success.

The measurement of disease progression and treatment response in clinical trials is particularly challenging in neurological diseases, especially in rare diseases where efficiency in clinical development is of paramount importance. Integration of ML methods into natural history models for rare diseases should be considered as a component of the toolkit to characterize sources of variability in disease trajectory. Well annotated and curated RWD sources are indispensable enablers. A successful example of such an application has been described in Huntington's disease (HD) using data from the Enroll-HD registry.[5] ML techniques, including k-means clustering and the gradient boosting tree-based ensemble ML algorithm XGboost, were applied to analyze clinical and functional measures over time in 4,961 individuals with HD. The analysis grouped patients into three clusters of progression: rapid, moderate, and slow progressors. Notably, the most predictive feature for cluster assignment was the cytosine adenine guanine-age product score, which is directly reflective of the molecular hallmark of HD, followed by factors such as years since symptom onset, a medical history of apathy, body mass index at enrollment, and age at enrollment. These findings provide valuable insights into the factors influencing the rate of disease progression in HD. We envision opportunities for such models to serve as Bayesian priors for enhancing the design and analysis of POC studies. The discovery of digital biomarker end points from longitudinal high-dimensional data collected with wearable devices is another emerging area of application for ML models. In one example, real-world movement data were collected from 21 patients with Duchenne muscular dystrophy (DMD) and 17 age-matched controls using a 17-sensor bodysuit over 12 months.[6] Novel movement behavioral patterns characterizing patients with DMD were identified, with AI/ML algorithms enabling the discovery of a movement behavioral biomarker (KineDMD). This novel ethomic biomarker offered better longitudinal disease predictions, reduced variability, and physiologically contextualizable age-related progression dynamics than traditional clinical assessments, such as the standard 6-minute walk distance test, raising valuable opportunities for more efficient clinical trial design for evaluating investigational therapeutics for DMD.

15326535, 2024, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.3153, Wiley Online Library on [02/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



15326535, 2024, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.3153, Wiley Online Library on [02/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 2 Explainable ML prediction of disease activity in patients with Multiple Sclerosis from cladribine trials. (**a**) Data from over 6 years of observation of 1,935 patients in the 3 phase III pivotal trials of cladribine (CLARITY, CLARITY-EXT, and ORACLE-MS) were assessed to predict whether a patient will have disease activity. (**b**) A supervised multivariate ML method was trained to take 57 input covariates for predicting future status of disease activity, and then tested on the left-out dataset. (**c**) SHAP explained the model prediction for each patient. The global view of the top 20 predictive covariates shows covariates contributions to discriminate between the patients who in future will have disease activity and those who do not. DA, disease activity; ML, machine learning; SHAP, SHapley Additive exPlanations.

Digital twins have been proposed as a way to streamline the drug development process and enhance efficiency in clinical trial design by enabling highly powered studies with smaller control groups. By leveraging generative AI, digital twins provide virtual representation or digital replica of a subject by modeling multimodal, high-dimensional, and longitudinal individual data, and enable simulations of how a patient's disease trajectory may evolve over time, factoring in variations if they were to receive either a placebo or standard of care, as illustrated in a seminal example in Alzheimer's disease.[7] TwinRCTs, a pioneering concept proposed by Unlearn.AI, leverage regulatory-qualified methods to seamlessly integrate patients' digital twins into clinical trials.[8] Crucially, TwinRCTs uses a regulatory-qualified PROCOVA framework to estimate treatment effects accurately. Patients are randomized to treatment or control groups, and at the trial's conclusion, information from their digital twins is harnessed. A prognostic model, developed based on historical data, guides the estimation of treatment effects, adjusting for patients' predicted outcomes (prognostic scores).[9] This procedure minimizes bias and optimizes efficiency, supporting its superiority over conventional covariate adjustment methods. The European Medicine Agency's (EMA) qualification of Unlearn's procedure in September 2022 further supports the application of digital twins in clinical trials.[10] Nevertheless, the implementation of digital twins in clinical trials is not without its challenges. Validation of digital twins and simulation models is a complex endeavor, particularly as alternate scenarios, not based on retrospective longitudinal data, are often absent from the ground truth. Despite these challenges, the integration of digital twins in clinical trials holds the promise of reducing sample sizes by 10 to 30%, thus increasing the efficiency of drug development while maintaining patient safety and regulatory compliance.

IDENTIFYING PREDICTORS OF TREATMENT OUTCOMES

An important application of ML models is in the identification of predictors of treatment response. In a recent example from our research group, disease activity in patients with multiple sclerosis (MS) was predicted using supervised multivariable explainable ML models of longitudinal data in the phase III clinical trials of cladribine.[11] Using XGboost, the models combined patient baseline characteristics, longitudinal magnetic resonance imaging data, neurological assessments, and laboratory measures in 1,935 patients to predict disease activity 3 and 6 months before it was clinically observed. Several predictive factors for early disease activity were identified, including the duration of cladribine treatment, the number of new combined unique active lesions, the number of new T1 hypointense black holes, and the age-related MS severity score. Difficulties in interpreting complex ML models, such as XGBoost and their predictions, sometimes limit the practical applicability of and confidence in ML. Explainable ML is aimed at overcoming these hurdles and enhancing uptake of ML model-informed findings in clinical settings. To this end, SHapley Additive exPlanations (SHAP) methods were incorporated. SHAP methods explain the model prediction for each patient by computing the contribution of each covariate to the prediction using game theory. The global ranking of important covariates is obtained by ordering the covariates by their mean absolute SHAP values for all patients. Importantly, graphical summarization of the results of SHAP analyses (**Figure 2**) permitted understandable communication of findings to non-technical audiences – a critical competency for maximizing the return on investment in advanced analytical models. Whereas the above example addressed disease activity and treatment response, these ML methods are equally valuable for elucidating and appreciating the relative importance of predictors of adverse effects of drug treatments and inform development of risk mitigation strategies. In one example, XGBoost ML models coupled with SHAP analyses helped to identify predictors of hyperglycemia during treatment with the AKT inhibitor anticancer agent ipatasertib.[12]

In oncology drug development, a key challenge, particularly in the development of immunotherapeutic agents, is the uncertainty in linkage between short-term response and long-term survival outcomes. Heterogeneity in molecular and radiographic response to anticancer agents is substantial, with opportunities to characterize this variability and relationships to overall survival using ML methods. Although total tumor size, assessed as the sum of longest diameters of target lesions, remains the primary metric used to assess radiographic response in oncology drug development, interlesion heterogeneity in tumor dynamics is seldom characterized, with often underappreciated implications for prediction of long-term outcomes. In investigations from our research team in metastatic colorectal cancer, an ML-based classification clustering of individual lesions (CICIL) methodology was developed to quantify heterogeneity in trajectory of individual target lesions.[13] When applied to 6,369 individual target lesions data from 1,781 patients with metastatic colorectal cancer receiving cetuximab-based therapy, CICIL-based metrics of heterogeneity in lesion-level tumor dynamics were found to be associated with the tumor molecular profile (KRAS mutation status) and identified as an important predictor of overall survival.[14,15]

Advances in liquid biopsy are enabling deep molecular characterization of cancer evolution and molecular response dynamics via circulating tumor DNA (ctDNA) measurements in clinical trials, whereas advances in radiomics have exponentially increased the information content in radiographic imaging data. Although linkage between changes in ctDNA or radiomic signatures and survival outcomes have been described[16] and progress has been made in applying pharmacometric methodologies to modeling tumor size

15326535, 2024, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.3153, Wiley Online Library on [02/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with integration of ctDNA and radiomics-based covariates,[17,18] disease progression modeling of longitudinal high-dimensional ctDNA or radiomics profiles for elucidating POC and for identifying patient subgroups with differential treatment response remain largely untapped opportunities in drug development settings. In a study involving 466 patients with non-small cell lung cancer from a clinical trial that evaluated atezolizumab-based treatments, ML was applied to jointly model the dynamics of multiple ctDNA features as predictors of overall survival.[19] Patients were stratified into high-risk and low-intermediate-risk groups, with significant differences in survival. In addition to the high prognostic value of baseline ctDNA levels, correlations of on-treatment ctDNA changes with treatment response suggested opportunities to improve patient risk stratification, when combined with radiographic imaging. Examples like these highlight the opportunity for model-based analyses assisted by ML methods to enhance the value of longitudinal liquid biopsy measurements in anticancer drug development. This is further reinforced by the recent US Food and Drug Administration (FDA) Draft Guidance on ctDNA measurements in early oncology drug development.[20] Envisioned applications include informing confidence in therapeutic potential, Go/No Go decision making, patient selection, and dosage optimization based on dose/schedule-related dynamics of disease- and mechanism-relevant ctDNA signatures in early oncology clinical development.

ENABLING INCLUSIVE CLINICAL DEVELOPMENT

Generalizability of findings in clinical trials regarding the benefit/ risk of investigational therapeutics to intended real world settings of clinical use requires adequately inclusive enrollment. This should ideally ensure that the intrinsic and extrinsic factors distributions in the clinical trial sample approximate those expected in the target patient population, while ensuring rational eligibility criteria that protect the safety of clinical trial participants. Efforts to broaden eligibility criteria in clinical trials are on the rise, recognizing that the rationale for exclusion criteria may not always be scientifically justified. ML-based analyses of clinical trial databases have informed forecasting the impact of eligibility criteria on demographic distributions (e.g., age and race) in a clinical trial sample. In one example, ML analyses evaluated the impact of 150 study traits generated from 14 Alzheimer's disease clinical trials on 4 FDA-approved treatments. At least 30 study traits were identified, which when used as exclusion criteria were shown to lower the age of the study population, thereby excluding older patients, leading to their under-representation in Alzheimer's disease clinical trials.[21]

Mining RWD in electronic health records (EHRs) using AI/ ML methods enables identification of patient factors associated with safety adverse events (AEs) and assessment of the impact of commonly applied exclusion criteria on patient outcomes, informing development of strategies to appropriately broaden eligibility criteria and/ or inform risk management. One such study was motivated by the observation of significant differences in eligibility criteria among advanced non-small cell lung cancer trials, even when they involved similar checkpoint inhibitor treatments. Through AI-based analyses of data from the Flatiron Health EHR database implemented in a computational framework *Trial PathFinder* that was developed for this purpose, it was possible to conclude that many commonly used laboratory-based exclusion criteria did not have a substantial impact on survival outcomes. Simulations of clinical trials with more inclusive criteria showed that broadening eligibility could have nearly doubled the pool of eligible patients, including more women and participants over 75 years old, without affecting the survival benefit or treatment withdrawals due to adverse effects.[22] In another example, an ML approach was developed to identify patient subgroups more or less likely to experience post-treatment severe adverse events using RWD in the OneFlorida+ Clinical Research Consortium, which contains longitudinal and linked patient-level RWD including data from Medicaid and Medicare claims, cancer registries, vital statistics, and EHRs from its clinical partners.[23] To account for the high dimensionality of RWD, a novel supervised topic modeling approach was developed that uses the severe AE (SAE) information as a weak supervision. The developed algorithm identified a set of clinical topics and derived novel patient representations based on them with reduced dimensionality (i.e., from thousands of clinical features to 40 topics), such that the patient subgroups with or without SAEs could be well separated with these representations. These examples illustrate the potential of ML approaches to understand factors governing patient outcomes in real-world practice settings, thereby providing valuable input to enhance diversity and inclusivity in clinical trials without compromising the balance of benefits and risks.

Global access to innovative medicines relies on an inclusive approach to drug development, enabled by regulatory frameworks such as the International Conference on Harmonization (ICH) E17 guideline that inform the design and analysis of multi-regional clinical trials (MRCTs). A key guiding principle of ICH E17 is that drug- and disease-related intrinsic and extrinsic factors should be characterized early to inform MRCT design.[24] Furthermore, efficiencies in trial design and analysis, including prespecified pooling of geographic regions or countries with similar drug- and disease-related intrinsic and extrinsic factors, is an opportunity offered by ICH E17. The question of similarity in disease-related intrinsic and extrinsic factors requires a robust characterization of sources of variability in disease trajectory/ progression. Disease progression modeling is an established approach in pharmacometrics to answer this question. Relevant data sources include adequately annotated RWD, patient-level data in pre-competitive collaborative consortia, and internal data from failed and successful clinical trials in the target indication. In recent examples from our research group in SLE[25] and the maintenance treatment of gastric cancer,[26] ML algorithms were embedded into pharmacometric workflows to inform covariate modeling in disease progression modeling. An objective of the covariate analysis was to understand the conservation of disease trajectory in Asian patient populations. In both examples, whereas a number of covariates were identified, Asian populations were shown to have disease trajectories that were generally consistent with non-Asian populations. Findings of this nature lend valuable support to the design of Asia-inclusive MRCTs. Importantly, when intrinsic or extrinsic sources of variability are identified in disease trajectory and their distribution is known or expected to display regional variation based on epidemiological

15326535, 2024, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.3153, Wiley Online Library on [02/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

research, appropriate control of such variability (e.g., through selection of appropriate stratification factors) will help mitigate the impact of heterogeneity, while enabling an inclusive approach to the design of MRCTs. To this end, the intersection of pharmacometric and ML-based covariate modeling methods is an ongoing area of research where ML-based workflows hold promise to enhance efficiency of covariate analyses in drug-disease modeling efforts.[27]

CONCLUDING REMARKS

Advanced analytical approaches, including but not limited to AI/ML, are increasingly enabling integrated consideration of multimodal, multisource, heterogenous sources of variability in modeling disease trajectory and the pharmacological response to treatments across diverse therapeutic areas. Such drug-disease knowledge management is vital in drug development for informing selection of the right patient population, optimizing clinical measurements of treatment response, enabling inclusive and efficient clinical trial designs, and ultimately driving efficient end-to-end evidence generation in pharmaceutical research and development. Advancing progress in this emerging area relies on interdisciplinary collaborations, including but not limited to disease area experts, clinical pharmacologists, pharmacometricians, biostatisticians, epidemiologists, and biomarker and data scientists, with commitment to continuous learning and a growth mindset. Of note, in the realm of AI/ML, it is essential to exercise caution to prevent biased and non-generalizable outcomes from data. Best practices should be followed by having human-in-the-loop and ensuring that the context of use and risk-informed credibility assessments are considered in the model development and validation.[2] With the rapid permeation of AI/ML into biomedical sciences and pharmaceutical research and development and the enthusiasm these advanced analytical techniques have gained in recent years, it is important to ensure purpose-oriented application to maximize impact.[28] Although we have highlighted the value of incorporating AI/ML algorithms in drug-disease modeling in this mini-review, not all questions and contexts of use will require these approaches. For example, a more traditional disease progression modeling analysis using established methods of covariate analysis in nonlinear mixed effects modeling will be well-suited in cases where dimensionality in the covariate space is limited. To this end, the scope of the problem and context of use should dictate the chosen methodology aligned with the principles applicable for credible practice of modeling and simulation applications in healthcare.[29] Explainable ML approaches are particularly valuable in this regard to enhance communication, interpretability, stakeholder understanding, and trust, which are crucial enablers for enhancing influence and impact.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

N.T. is an employee of Ares Trading S.A. (an affiliate of Merck KGaA, Darmstadt, Germany), Lausanne, Switzerland. K.V. is an employee of EMD Serono, Research & Development Institute, Inc., a business of Merck KGaA. As an Associate Editor for Clinical Pharmacology & Therapeutics, Karthik Venkatakrishnan was not involved in the review or decision process for this paper.

© 2023 Merck KGaA Darmstadt. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals LLC 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. Venkatakrishnan, K. & Benincosa, L.J. Diversity and inclusion in drug development: rethinking intrinsic and extrinsic factors with patient centricity. *Clin. Pharmacol. Ther.* **112**, 204–207 (2022).
2. Terranova, N. *et al*. Artificial intelligence for quantitative modeling in drug discovery and development: an innovation and quality consortium perspective on use cases and best practices. *Clin. Pharmacol. Ther.* (2023). https://doi.org/10.1002/cpt.3053. Online ahead of print.
3. Terranova, N., Venkatakrishnan, K. & Benincosa, L.J. Application of machine learning in translational medicine: current status and future opportunities. *AAPS J.* **23**, 74 (2021).
4. Choi, M.Y. *et al*. Machine learning identifies clusters of longitudinal autoantibody profiles predictive of systemic lupus erythematosus disease outcomes. *Ann. Rheum. Dis.* **82**, 927–936 (2023).
5. Ko, J. *et al*. Clustering and prediction of disease progression trajectories in Huntington's disease: an analysis of enroll-HD data using a machine learning approach. *Front. Neurol.* **13**, 1034269 (2023).
6. Ricotti, V. *et al*. Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. *Nat. Med.* **29**, 95–103 (2023).
7. Fisher, C.K., Smith, A.M. & Walsh, J.R. Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci. Rep.* **9**, 13622 (2019).
8. <https://www.unlearn.ai/twinrcts>
9. Schuler, A., Walsh, D., Hall, D., Walsh, J., Fisher, C. & Critical Path for Alzheimer's Disease; Alzheimer's Disease Neuroimaging Initiative; Alzheimer's Disease Cooperative Study. Increasing the efficiency of randomized trial estimates via linear adjustment for a prognostic score. *Int. J. Biostat.* **18**, 329–356 (2021).
10. European Medical Agency, Committee for Medicinal Products for Human Use (CHMP). Qualification opinion for prognostic covariate adjustment (PROCOVA™). EMADOC-1700519818-907465 (2022). <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-prognostic-covariate-adjustment-procovatm_en.pdf>
11. Basu, S., Munafo, A., Ben-Amor, A.F., Roy, S., Girard, P. & Terranova, N. Predicting disease activity in patients with multiple sclerosis: an explainable machine-learning approach in the Mavenclad trials. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 843–853 (2022).
12. Harun, R., Sane, R., Yoshida, K., Sutaria, D.S., Jin, J.Y. & Lu, J. Risk factors of hyperglycemia after treatment with the AKT inhibitor Ipatasertib in the prostate cancer setting: a machine learning–based investigation. *JCO Clin. Cancer Inform.* **7**, e2200168 (2023).
13. Terranova, N., Girard, P., Ioannou, K., Klinkhardt, U. & Munafo, A. Assessing similarity among individual tumor size lesion dynamics: the CICIL methodology. *CPT Pharmacometrics Syst. Pharmacol.* **7**, 228–236 (2018).
14. Vera-Yunca, D., Girard, P., Parra-Guillen, Z.P., Munafo, A., Trocóniz, I.F. & Terranova, N. Machine learning analysis of individual tumor lesions in four metastatic colorectal cancer clinical studies: linking tumor heterogeneity to overall survival. *AAPS J.* **22**, 1–12 (2020).
15. Vera-Yunca, D., Parra-Guillen, Z.P., Girard, P., Trocóniz, I.F. & Terranova, N. Relevance of primary lesion location, tumour heterogeneity and genetic mutation demonstrated through tumour growth inhibition and overall survival modelling in metastatic colorectal cancer. *Br. J. Clin. Pharmacol.* **88**, 166–177 (2022).
16. Vega, D.M. *et al*. Changes in circulating tumor DNA reflect clinical benefit across multiple studies of patients with non–small-cell lung cancer treated with immune checkpoint inhibitors. *JCO Precis. Oncol.* **6**, e2100372 (2022).

15326535, 2024, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.3153, Wiley Online Library on [02/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

17. Janssen, J.M. *et al*. Longitudinal nonlinear mixed effects modeling of EGFR mutations in ctDNA as predictor of disease progression in treatment of EGFR-mutant non-small cell lung cancer. *Clin. Transl. Sci.* **15**, 1916–1925 (2022).
18. Courlet, P. *et al*. Modeling tumor size dynamics based on real-world electronic health records and image data in advanced melanoma patients receiving immunotherapy. *CPT Pharmacomet. Syst. Pharmacol.* **12**, 1170–1181 (2023).
19. Assaf, Z.J.F. *et al*. A longitudinal circulating tumor DNA-based model associated with survival in metastatic non-small-cell lung cancer. *Nat. Med.* **29**, 859–868 (2023).
20. U.S. Department of Health and Human Services Food and Drug Administration. Use of circulating tumor DNA for early-stage solid tumor drug development guidance for industry (2022). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-circulating-tumor-deoxyribonucleic-acid-early-stage-solid-tumor-drug-development-draft-guidance>
21. Chen, A. *et al*. Impacts of eligibility criteria on trial participants' Age in Alzheimer's disease clinical trials. *AMIA Annu. Symp. Proc.* **2022**, 368–376 (2022). American Medical Informatics Association.
22. Liu, R. *et al*. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* **592**, 629–633 (2021).
23. Xu, J., Zhang, H., Zhang, H., Bian, J. & Wang, F. Machine learning enabled subgroup analysis with real-world data to inform clinical trial eligibility criteria design. *Sci. Rep.* **13**, 613 (2023).
24. Venkatakrishnan, K. *et al*. Asia-inclusive clinical Research and Development enabled by translational science and quantitative clinical pharmacology: toward a culture that challenges the status quo. *Clin. Pharmacol. Ther.* **113**, 298–309 (2023).
25. Goteti, K. *et al*. Disease trajectory of SLE clinical endpoints and covariates affecting disease severity and probability of response: analysis of pooled patient-level placebo (standard-of-care) data to enable model-informed drug development. *CPT Pharmacometrics. Syst. Pharmacol.* **12**, 180–195 (2023).
26. Terranova, N. *et al*. Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN gastric 100 phase III trial of avelumab. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 333–347 (2022). https://doi.org/10.1002/psp4.12754
27. Sibieude, E., Khandelwal, A., Hesthaven, J.S., Girard, P. & Terranova, N. Fast screening of covariates in population models empowered by machine learning. *J. Pharmacokinet. Pharmacodyn.* **48**, 597–609 (2021).
28. Shahin, M.H. *et al*. Artificial intelligence: from buzzword to useful tool in clinical pharmacology. *Clin. Pharmacol. Ther.* (2023). https://doi.org/10.1002/cpt.3083. Online ahead of print.
29. Erdemir, A. *et al*. Credible practice of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective. *J. Transl. Med.* **18**, 369 (2020).

15326535, 2024, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.3153, Wiley Online Library on [02/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License