

Data Science in Clinical Pharmacology and Drug Development for Improving Health Outcomes in Patients

Richard W. Peck^{1,*†}, Pratik Shah^{2,†}, Spiros Vamvakas³ and Piet H. van der Graaf^{4,5}

Data Science is about drawing useful conclusions from large and diverse data sets through exploration, prediction, and inference.¹ Alternatively, Wikipedia defines data science as “a multidisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.”² Clinical pharmacology has always been a quantitative discipline making extensive use of different analytical methods to extract knowledge from data, and the rapidly evolving field of data science offers many opportunities for clinical pharmacologists. In this issue of *Clinical Pharmacology & Therapeutics (CPT)*, we explore many of these opportunities and how they can benefit this discipline. This issue is also the first to be fully dedicated to a single topic, where the perspectives, reviews, and the original research articles are all related to the same theme. We also introduce the first tutorials, a new article type for *CPT*, dedicated to helping readers to understand new areas of science. Data science is impacting the data sources and types that are becoming available, how data are captured and made available to users, and how they are analyzed and interpreted.

This issue considers how each of these can be relevant to clinical pharmacologists and clinical development researchers (**Figure 1**).

Traditionally, clinical pharmacology, and indeed much of drug discovery and development, has proceeded through carefully designed experiments and clinical trials that test a hypothesis. Phase III clinical outcome trials evaluating new therapies and vaccines are among the most complex experiments performed in medicine, and a common theme is the difficulty of predicting clinical results in a wider patient base after regulatory approvals. The high cost of clinical trials, low success rates, and potentially reduced efficacy of approved therapies in larger populations can cost healthcare industries, government, and academic research hospitals millions of dollars each year, may drive up costs and delay life-saving treatments to patients, and in some cases lead to adverse events. Data in clinical trials are usually captured to answer the specific question under investigation. Real-world data (RWD) is data collected outside the boundaries of a specific experiment or clinical trial, often for reasons other than scientific hypothesis testing, and is thought of as an important source of additional or novel information

¹Pharma Research and Development, Roche Innovation Center Basel, Basel, Switzerland; ²Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; ³European Medicines Agency, Amsterdam, The Netherlands; ⁴Certara, Canterbury, UK; ⁵University of Leiden, Leiden, The Netherlands.

*Correspondence: Richard W. Peck (richard.peck@roche.com)

†Equal contribution.

Received January 30, 2020; accepted January 30, 2020. doi:10.1002/cpt.1803



Figure 1 *Clinical Pharmacology & Therapeutics* April 2020 cover image.

often not captured during clinical trials. A key challenge of using RWD as evidence for clinical development or efficacy, though, is the lack of randomization often resulting in limited causal inference in observational data and interference by confounding factors. However, data providers are now making such data available for scientific investigation, leading several data scientists to incorporate emerging analytics, including deep learning for generating actionable evidence to overcome key limitations of the current drug development process.³ The US Food and Drug Administration (FDA) has also published draft white papers and guidelines to incorporate RWD in clinical decision making.⁴

A recent survey paper showed that availability of observational data is quite limited, their use is limited for clinical development, and randomized controlled trial results continue to dominate regulatory submissions.⁵ However, the utility of RWD should not be constrained to accelerate clinical development, but rather expanded to isolate, enrich, and study responses, outcomes, and adverse events in more diverse populations reflecting heterogeneity in genotypes, genders, and socio-economic biases which have profound impact on therapeutic efficacy and generalization of clinical trial results. Thus, there is a significant opportunity to provide societal impact and benefit by conducting carefully designed and hypothesis-driven RWD studies. There are also many opportunities where RWD can benefit

clinical pharmacologists, including streamlining or even replacing clinical trials in a few instances,^{6,7} informing on difficult-to-study populations, such as children,⁷ or rare diseases, drug repurposing,⁸ pharmacovigilance,⁹ pharmacokinetic/pharmacodynamic modeling,¹⁰ and physiologically-based pharmacokinetic modeling.¹¹ Given the range of opportunities, we must also understand the issues and challenges in how RWD should be used¹²⁻¹⁴ and ensure that appropriate validation is undertaken for new methods.¹⁵

RWD is also used commonly in the context of the explosion of data available from -omics, continuous, ambulatory (usually in the real world) patient-monitoring technology, including wearables and other high-capacity data capture and analytical methods (often referred to as Big Data). The potential of Big Data in drug development is of interest to regulators,¹⁵ and clinical pharmacologists are well placed to guide integration of Big Data analyses with modeling¹⁶ as a way to use, analyze, and interpret the data. One of the new tutorial articles in this issue describes how to combine -omics data with mechanistic modeling, a core competency of our discipline.¹⁷ Here, too, it is important to understand the limitations. We must be familiar with how data are collected using novel technologies, such as handheld devices or wearables,^{18,19} and contribute to the development of appropriate reporting standards to ensure data quality.²⁰

Machine and deep learning, whether directly through interacting with smart phones, smart speakers, and lifestyle apps, or indirectly, for example through life insurance companies predicting our risk and setting our premiums or banks making investment decisions with our savings, is now being routinely used. In most areas of life sciences and health care there is widespread consideration of the potential opportunity for and impact of machine learning and artificial intelligence (AI), including deep learning. The vast majority of applications of deep learning in health care today are examples of pattern recognition, such as assessment of diagnostic images, which in computer science is considered narrow AI. These efforts are noteworthy and exciting but do not represent fundamental research in AI systems capable of cognition, reasoning, generative tasks, reinforcement learning, and other human-like traits, referred to as "general" or "broad AI," which will also have significant impact on clinical data science in coming years. A recent description of generative deep learning algorithms for image translation and computational hematoxylin and eosin staining of whole slide pathology images is an example of expanding pattern recognition to generative tasks for faster clinical inference.²¹ There is also promise for using machine and deep learning to find simple patterns in patient data that could provide early diagnostic or prognostic tests or guide therapeutic decision making. Early examples of AI agents exhibiting human learning capabilities, such as reinforcement driven by actions, rewards, and penalties have also been described recently for dose deescalation algorithms using pharmacokinetic/pharmacodynamic models.²²

Thus, the terms "artificial intelligence" and "machine learning" are widely used, including in Hollywood film titles, but one needs to interpret them fairly and accurately. The second tutorial in this issue provides a primer in basic machine learning to demystify some concepts, methods, and applications.²³ In our view, there are many exciting opportunities for machine learning to help clinical pharmacologists and drug discovery and development.^{24,25} Often the biggest advances occur when different disciplines intersect, and pharmacometrics should be fertile ground to benefit from machine learning—indeed there is already quite a literature on this topic. In this issue we add to the discussion with examples of how machine

learning can help pharmacometricians,²⁶ precision medicine and precision dosing,^{25,27} identification of useful drug combinations,²⁸ and pharmacovigilance. We are happy to agree with Brian Corrigan²⁹ that AI will not replace clinical pharmacologists but instead will help us be more successful.

One way to make data "bigger" is open sharing such that all of us have easier access to more data than one can collect in their own work. In drug discovery and development, as with any activity incentivized and rewarded through intellectual property, some data will have to remain confidential for a period of time, but there is much else that can be shared immediately, and we strongly encourage such activities.^{30,31} As methods to investigate and learn from Big Data advance, this sharing becomes essential to allow us to tackle big problems. Examples in this issue include improving pediatric prescribing through collaborating to understand development ontogeny³² and the identification of novel treatments for devastating, untreatable, or almost untreatable diseases.³³ Sharing models, or at least making them transparent, is equally important as a means to stimulate understanding, to further advances, and to encourage wider uptake of the models themselves.³⁰

© 2020 The Authors *Clinical Pharmacology & Therapeutics* © 2020 American Society for Clinical Pharmacology and Therapeutics

1. Adhikari, A. & DeNero, J. *The Foundations of Data Science* (UC Berkeley, Berkeley, CA, 2018) <<https://www.inferentialthinking.com/chapters/intro.html>>.
2. Wikipedia: Data science <https://en.wikipedia.org/wiki/Data_science>. Accessed January 16, 2019.
3. Shah, P. *et al.* Artificial intelligence and machine learning in clinical development: a translational perspective. *npj Digital Med* **2**, 69 (2019).
4. US Food and Drug Administration. Guidance for industry and Food and Drug Administration staff: use of real-world evidence to support regulatory decision-making for medical devices <<https://www.fda.gov/downloads/medicaldevices/deviceeregulationandguidance/guidancedocuments/ucm513027.pdf>> (2017). Accessed August 31, 2019.
5. Bartlett, V.L., Dhruva, S.S., Shah, N.D., Ryan, P. & Ross, J.S. Feasibility of using real-world data to replicate clinical trial evidence. *JAMA Netw Open* **2**, e1912869 (2019).
6. Schmidli, H., Häring, D.A., Thomas, M., Cassidy, A., Weber, S. & Bretz, F. Beyond randomized clinical trials: use of external

- controls. *Clin. Pharmacol. Ther.* **107**, 806–816 (2020).
7. Goulooze, S.C. *et al.* Beyond the randomized clinical trial: innovative data science to close the pediatric evidence gap. *Clin. Pharmacol. Ther.* **107**, 786–795 (2020).
 8. Xu, H., Li, J., Jiang, X. & Chen, Q. Electronic health records for drug repurposing: current status, challenges, and future directions. *Clin. Pharmacol. Ther.* **107**, 712–714 (2020).
 9. Candore, G., Hedenmalm, K., Slattery, J., Cave, A., Kurz, X. & Arlett, P. Can we rely on results from IQVIA Medical Research Data UK converted to the observational medical outcome partnership common data model? *Clin. Pharmacol. Ther.* **107**, 915–925 (2020).
 10. Choi, L. *et al.* Development of a system for post-marketing population pharmacokinetic and pharmacodynamic studies using real-world data from electronic health records. *Clin. Pharmacol. Ther.* **107**, 934–943 (2020).
 11. Wang, K. *et al.* Real-World data and physiologically-based mechanistic models for precision medicine. *Clin. Pharmacol. Ther.* **107**, 694–696 (2020).
 12. McDonough, C.W. Translational informatics connects real-world information to knowledge in an increasingly data-driven world. *Clin. Pharmacol. Ther.* **107**, 738–741 (2020).
 13. Franklin, J.M., Glynn, R.J., Suissa, S. & Schneeweiss, S. Emulation differences vs. biases when calibrating Real-World Evidence findings against Randomized Controlled Trials. *Clin. Pharmacol. Ther.* **107**, 735–737 (2020).
 14. Keizer, R.J. *et al.* Get real: integration of real-world data to improve patient care. *Clin. Pharmacol. Ther.* **107**, 722–725 (2020).
 15. Cave, A., Brun, N.C., Sweeney, F., Rasi, G. & Senderovitz, T. Big data – how to realize the promise. *Clin. Pharmacol. Ther.* **107**, 753–761 (2020).
 16. Mistry, H.B. & Orrell, D. Small models for big data. *Clin. Pharmacol. Ther.* **107**, 710–711 (2020).
 17. Lazarou, G., Chelliah, V., Small, B.G., Walker, M., van der Graaf, P.H. & Kierzek, A.M. Integration of omics data sources to inform mechanistic modeling of immunoncology therapies: a tutorial for clinical pharmacologists. *Clin. Pharmacol. Ther.* **107**, 858–870 (2020).
 18. Pradhan, V., Heatherington, A.C. & Bhattacharya, I. End points, collection, processing, and time: four key elements to consider when planning for use of handheld devices in a drug development setting. *Clin. Pharmacol. Ther.* **107**, 715–718 (2020).
 19. Omberg, L., Neto, E.C. & Mangravite, L.M. Data science approaches for effective use of mobile device–based collection of real-world data. *Clin. Pharmacol. Ther.* **107**, 719–721 (2020).
 20. Febbo, P.G. *et al.* Minimum technical data elements for liquid biopsy data submitted to public databases. *Clin. Pharmacol. Ther.* **107**, 730–734 (2020).
 21. Rana, A. *et al.* High accuracy tumor diagnoses and benchmarking of hematoxylin and eosin stained prostate core biopsy images generated by explainable deep neural networks <<https://arxiv.org/abs/1908.01593>> (2019).
 22. Yauney, G. & Shah, P. Reinforcement learning with action-derived rewards for chemotherapy and clinical trial dosing regimen selection. *Proc. Mach. Learn. Res.* **85**, 161–226 (2018).
 23. Badillo, S. *et al.* An introduction to machine learning. *Clin. Pharmacol. Ther.* **107**, 871–885 (2020).
 24. Zhavoronkov, A., Vanhaelen, Q. & Oprea, T.I. Will artificial intelligence for drug discovery impact clinical pharmacology? *Clin. Pharmacol. Ther.* **107**, 780–785 (2020).
 25. Liu, Q. *et al.* Application of machine learning in drug development and regulation: current status and future potential. *Clin. Pharmacol. Ther.* **107**, 726–729 (2020).
 26. Koch, G., Pfister, M., Daunhawer, I., Wilbaux, M., Wellmann, S. & Vogt, J.E. Pharmacometrics and machine learning partner to advance clinical data analysis. *Clin. Pharmacol. Ther.* **107**, 926–933 (2020).
 27. Ribba, B., Dudal, S., Lavé, T. & Peck, R.W. Model-informed artificial intelligence: reinforcement learning for precision dosing. *Clin. Pharmacol. Ther.* **107**, 853–857 (2020).
 28. Wang, Z., Li, H. & Guan, Y. Machine learning for cancer drug combination. *Clin. Pharmacol. Ther.* **107**, 749–752 (2020).
 29. Corrigan, B.W. Artificial intelligence and machine learning: will clinical pharmacologists be needed in the next decade? The John Henry question. *Clin. Pharmacol. Ther.* **107**, 697–699 (2020).
 30. Kirouac, D.C. Open models for clinical pharmacology. *Clin. Pharmacol. Ther.* **107**, 700–702 (2020).
 31. Burton, J., Bhattacharya, S., Romero, K. & Conrado, D.J. Open data for clinical pharmacology. *Clin. Pharmacol. Ther.* **107**, 703–706 (2020).
 32. Burckart, G.J. *et al.* Scientific and regulatory considerations for an ontogeny knowledge base for pediatric clinical pharmacology. *Clin. Pharmacol. Ther.* **107**, 707–709 (2020).
 33. Conrado, D.J. *et al.* Challenges in Alzheimer’s disease drug discovery and development: the role of modeling, simulation and open data. *Clin. Pharmacol. Ther.* **107**, 796–805 (2020).