frontiers | Frontiers in Pharmacology

TYPE Perspective
PUBLISHED 17 February 2023
DOI 10.3389/fphar.2022.1094281

Check for updates

OPEN ACCESS

EDITED BY
Zinnia P. Parra-Guillen,
University of Navarra, Spain

REVIEWED BY
Nadia Terranova,
Merck, Switzerland

*CORRESPONDENCE
Benjamin Ribba,
benjamin.ribba@roche.com

SPECIALTY SECTION
This article was submitted to Translational Pharmacology, a section of the journal Frontiers in Pharmacology

RECEIVED 09 November 2022
ACCEPTED 12 December 2022
PUBLISHED 17 February 2023

CITATION
Ribba B (2023), Reinforcement learning as an innovative model-based approach: Examples from precision dosing, digital health and computational psychiatry.
Front. Pharmacol. 13:1094281.
doi: 10.3389/fphar.2022.1094281

Reinforcement learning as an innovative model-based approach: Examples from precision dosing, digital health and computational psychiatry

Benjamin Ribba*

Roche Pharma Research and Early Development (pRED), F. Hoffmann-La Roche Ltd, Basel, Switzerland

Model-based approaches are instrumental for successful drug development and use. Anchored within pharmacological principles, through mathematical modeling they contribute to the quantification of drug response variability and enables precision dosing. Reinforcement learning (RL)—a set of computational methods addressing optimization problems as a continuous learning process—shows relevance for precision dosing with high flexibility for dosing rule adaptation and for coping with high dimensional efficacy and/or safety markers, constituting a relevant approach to take advantage of data from digital health technologies. RL can also support contributions to the successful development of digital health applications, recognized as key players of the future healthcare systems, in particular for reducing the burden of non-communicable diseases to society. RL is also pivotal in computational psychiatry—a way to characterize mental dysfunctions in terms of aberrant brain computations—and represents an innovative modeling approach forpsychiatric indications such as depression or substance abuse disorders for which digital therapeutics are foreseen as promising modalities.

KEYWORDS

pharmacometrics, digital health, reinforcement learning, precision dosing, computational psychiatry

1 Reinforcement learning for precision dosing

Precision dosing, or the ability to identify and deliver the right dose and schedule (i.e. the dose and schedule with highest likelihood of maximizing efficacy and minimizing toxicity), is critical for public health and society. Precision dosing is not only important for marketed drugs to reduce the consequences of imprecise dosing in terms of costs and adverse events; but also for therapeutics in development to reduce attrition, often related to the challenge of precisely characterizing the therapeutic window due to a suboptimal understanding of drug-response variability. Achieving the benefit to society of precision dosing requires the identification of the main drivers of response variability, as early as possible in the drug development process, and the deployment into clinical practice through an infrastructure designed for real-time dosing decisions in patients (Maxfield and Zineh, 2021; Peck, 2021).

Model-based approaches to clinical pharmacology, also known as clinical pharmacometrics (PMX) play a critical role in precision dosing. First, they contribute to the identification of the determinants of response variability through quantitative analysis of pharmacokinetic (PK) and pharmacodynamics (PD) relationships, and second, they constitute a central part of the infrastructure providing a simulation engine, predicting individual patient's response to a dose, and from which optimal dosing is identified through reverse engineering. Often this reverse engineering comprises two steps: first the PMX model's individual parameters are calculated through Bayesian inference, i.e. through the calculation of the mode of posterior distribution (maximum a posteriori or MAP); second, an optimal dosing scheduling is calculated, often *via* an heuristic approach through simulating various feasible dosing scenarios on inferred individuals model's instances.

Many examples exist in literature describing relevant PKPD models for precision dosing. For instance, in oncology, a model describing the time course of neutrophils following chemotherapy treatment is an ideal candidate for optimizing chemotherapy delivery (see (Friberg et al., 2002) as an example). Studies have also reported clinical investigations of model-based precision dosing approaches. For instance, the clinical study "MODEL1" was a phase I/II trial and a clear clinical attempt of a personalized dosing regimen of docetaxel and epirubicin patients with metastatic breast cancer and was shown to lead to improved efficacy-toxicity balance (Henin et al., 2016).

Reinforcement learning (RL) was also used for precision dosing. Still in oncology, Maier et al. extended the classical framework of model-driven precision dosing with RL coupled or not with data assimilation techniques (Maier et al., 2021). Previously, RL applications—although without clinical confirmation—were developed for brain tumors (Yauney and Shah, 2018) based on a model of tumor size response to chemotherapy (Ribba et al., 2012). We have recently evaluated the performance of RL algorithms for precision dosing of propofol for general anesthesia and for which a meta-analysis showed that the monitoring of the bispectral index (BIS)—a PD endpoint—contributes to reduce the amount of propofol given and the incidence of adverse reactions (Wang et al., 2021). In (Ribba et al., 2022), we performed a theoretical analysis of propofol precision dosing confronting RL to hallmarks of clinical pharmacology problems during drug development, i.e. the low number of patients and tested dosing regimen, the incomplete understanding of the drivers of response and the presence of high variability in the data.

While RL does not present as a universal solution for all types of precision dosing problems, it is an interesting modeling paradigm worth exploring. In comparison to the way PMX traditionally addresses precision dosing, RL presents several advantages. First, the possibility to take into account high dimensional PKPD variables while classical model-based approaches are often limited to a low number of variables (plasma concentration and one endpoint). In doing so, it represents an opportunity for the integration of digital health data such as from wearable devices or digital health technologies in general. Second, the definition of the precision dosing policy in a dynamic and adaptable manner through the continuous learning of the algorithm through real and simulated experience (data). RL is an approach by which both the underlying model and the optimal dosing rules are learnt simultaneously while for classical approaches, these represent two sequential steps: in other words, the consequence of the dose does not influence the model structure. Recently, studies have been published illustrating methodologies for adapting PKPD model structures through data assimilation (Lu et al., 2021; Bram et al., 2022). While high dosing frequency is not a prerequisite condition for the applicability of RL to precision dosing, this approach is well suited when the solution space of dosing is large, making heuristic approaches to find optimal dosing solutions inadequate. In our example on propofol, dosing could happen every 5 s so over a short period of 2 min, the space of solutions to explore when considering dichotomous dosing even is greater than 16 million possibilities.

RL is at the crossroads between two scientific fields. First, the field of learning by trial and error that started with the study of the psychology of animal learning and second, the field of optimal control (Sutton and Barto, 2018). RL are often formally described with Markov Decision Process or MDP which includes all important features a learning agent should have, namely, being able to sense the environment, being able to take action and have clarity on the goal. In RL, a learning agent takes an action and, as a result, transitions from one state to another. After each action taken, the interaction between the agent and its environment produces a reward. The goal of the RL problem is to map actions to situations (state), i.e. knowing which actions to take in each state to maximize the accumulated reward. As long as the optimization problem can be formulated within the MDP framework, RL can be applied and its efficiency explored.

For precision dosing of propofol, the state can be represented by a table, an approach also called tabular solution methods. In the next two sections, the state will be defined by a continuous function. The reward was determined based on the value reached by the BIS as a direct consequence of the action taken: the closer the BIS to the target, the higher the reward. Finally, given the theoretical study, the true PKPD model (linking the dose application to BIS) was used as an experience (data) generator. The left column of Table 1 summarizes the characteristics of the application of RL to the propofol precision dosing problem.

The minimal set of RL characteristics makes it a very flexible paradigm, suitable for a large variety of problems. Herein, we will in fact illustrate this flexibility by illustrating how this framework

TABLE 1 Main characteristics of RL algorithm implementation to the precision dosing of pharmacological interventions (left column); the precision dosing of digital intervention (middle column); and computational psychiatry (right column). While there are multiple similarities between the precision dosing of pharmacological and digital interventions, the application of RL in computational psychiatry shows as a paradigm shift. RL computational machinery is not deployed as a technical approach to address the optimal control problem of precision dosing but is fitted to (cognitive task) data assuming the algorithm itself presents mechanistic similarities with how brain's participants functioned during the task.

<table>
<tr><th></th><th>Precision dosing of a pharmacological intervention</th><th>Precision dosing of a digital intervention</th><th>Computation psychiatry</th></tr>
<tr><td>Study case [References]</td><td>Optimal dosing of propofol administration (Ribba et al., 2022)</td><td>Just-in-time-adaptive-intervention for HeartSteps, mobile app aimed at reducing physical inactivity (Liao et al., 2020)</td><td>Population analysis of signal-detection task in anhedonic subjects (Huys et al., 2013)</td></tr>
<tr><td>Type of RL solution</td><td>Tabular</td><td colspan="2">Continuous</td></tr>
<tr><td rowspan="2">State</td><td colspan="2"><i>Is directly linked to the state of the patient</i></td><td><i>Is linked to the situation the participant to the task is presented with and based on which an action must taken</i></td></tr>
<tr><td>PK drivers and/or PD endpoint such as the BIS</td><td>Contextual drivers (e.g. weather conditions, time of the day) and patient-related status derived from wearable device equipment</td><td>Belief of the correctness (weight) of each stimuli present in the task</td></tr>
<tr><td>Action</td><td>Dose or not</td><td>Dose (walking suggestion message) or not</td><td>Participant's answer choice</td></tr>
<tr><td rowspan="2">Reward</td><td colspan="2"><i>Defined to enable the algorithm converging to the optimal dosing solution</i></td><td><i>Corresponds to whether the answer is correct or wrong</i></td></tr>
<tr><td>Simple function of BIS leading to high reward when actual BIS is close to its target</td><td>Step count in the 30 min window after each decision time</td><td>Automatically derived from the answer as per task design and setup</td></tr>
<tr><td rowspan="2">Use of simulated experience?</td><td colspan="2"><i>Yes</i></td><td><i>No</i></td></tr>
<tr><td>The true underlying PKPD model is used</td><td>Linear model assimilating real data</td><td>No need for simulated experience, RL algorithm is mapped to the trial-by-trial data</td></tr>
<tr><td>Algorithm</td><td>Temporal difference Q-learning</td><td>Thomson Sampling</td><td>Temporal difference Q-learning</td></tr>
<tr><td rowspan="2">Free parameters</td><td colspan="2"><i>Used to calibrate model of patient's response to dosing event</i></td><td><i>Used to calibrate RL algorithm</i></td></tr>
<tr><td>Parameters of the PKPD model</td><td>Parameters of the linear model for reward prediction under alternative dosing scenarios</td><td>Learning rate and reward sensitivity parameter</td></tr>
</table>

can be viewed as a bridge between a priori distinct areas such as precision dosing of pharmacological drugs, digital health and computational psychiatry.

In the appendix, we propose to demystify how RL algorithms—such as temporal difference Q-learning, repeatedly mentioned here—work, taking a simple illustration from video gaming.

2 Reinforcement learning in digital health

For several years, many reports have indicated the key importance of digital health for reducing the burden to society of non-communicable diseases such as cardiovascular, diabetes, cancer or psychiatric diseases, in part due to the aging of the population and—paradoxically—the success of pharmacologically-based interventions in increasing life expectancy while being affected by pathological conditions (Fleisch et al., 2021). Prevention and interventions targeting lifestyle are essential tools to address this societal challenge of ever-growing importance as our healthcare systems risk collapse under cost pressure.

In 2008, it was estimated that physical inactivity causes 6% of the burden of coronary heart disease, 7% of type II diabetes, 10% of breast cancer and 10% of colon cancer and overall the cause of more than 5.3 million of the 57 million deaths which occurred that year (Lee et al., 2012). In that study, the authors also estimated that with 25% reduction of physical inactivity, 1.3 million of deaths could be averted every year. Given the constant increase of smartphone coverage worldwide, it is natural to think of mobile health technologies to support healthy lifestyle habits and prevention. The thinktank Metaforum from KU Leuven dedicated its position paper 17 on the use of wearables and mobile technologies for collecting information on individual behavior and physical status—combined with data from individual's environment—to personalize recommendations (interventions) bringing the subject to adopt a healthier lifestyle (Claes, 2022).

When the intervention is intended to have a therapeutic benefit, it falls in the field of digital therapeutics when associated with demonstration of clinical effectiveness and approved by regulatory bodies (Sverdlov et al., 2018). This point of junction between digital health applications and pharmacological drugs represents a ground for attempting to reframe PMX—a recognized key player in the development of the latter—as a key support to the development of the former, in particular when it comes to precision dosing for digital health.

The precision dosing of digital therapeutics overlaps with the concept of just-in-time adaptive intervention or JITAI (Nahum-Shani et al., 2018). In the mobile technology literature, JITAI has been primarily considered as a critical topic for increasing adherence and retention of users; but within a therapeutic perspective, it should encompass both the topic of adherence and retention to the therapeutic modality and the topic of its optimal dosing in order to maximize clinical benefit. For clarity, these two different learning problems should be distinguished as many existing applications focus primarily on the first one. For example, a growing number of mobile applications developed under the concept of virtual coaching aim to optimize the design of the interventions (time and content, e.g. messages sent by the app to the users with the form of a prompt appearing on a locked screen) to incite the user to take actions. HeartSteps was designed to encourage user to increase their physical activity and where content delivery, such as tailored walking suggestion messages, is optimized with an RL algorithm (Liao et al., 2020). Here, RL is used to address the first learning problem: How to deliver the content so that the user is doing what is recommended. We each need different forms of prompting and potentially different forms of exercise to increase our physical activity. Overall, this problem is similar to that of adherence to a pharmacological regimen. But a second problem is: what is the right dose of the desired intervention? In other words: How many steps is optimal for each patient? This is the usual precision dosing problem for drugs and there is a clear opportunity for digital health applications to extend the domain of application of JITAIs to that problem as well.

One of the particularly interesting aspects of the research on RL algorithms for HeartSteps is that, beyond the innovative nature of the work purely related to the design of personalized interventions, it also includes ways to objectively evaluate its efficiency. An experimental design called micro-randomized trial (MRT) is proposed as a framework to evaluate the effectiveness of personalized *versus* non-personalized interventions (Klasnja et al., 2015; Qian et al., 2022). The principle of MRT is to randomize the interventions multiple times for each subject. Statistical approaches have been studied to leverage MRT-derived data in order to inform treatment effects and the response variability (Qian et al., 2020). In the theoretical propofol example described in the previous section, we used the true PKPD model to simulate experience. In the real-life RL application of HeartSteps, the authors had the objective to design a method for learning quickly and for accommodating noisy data (Liao et al., 2020). To address these points, the authors used a simulation engine to enhance data collected from real experience and this simulation engine was built with simple linear models. Precisely, the authors modeled the difference in reward function under alternative dosing options with low dimensional linear models, which features were selected based on retrospective analysis of previous HeartSteps data and based on experts' guidance. The precision dosing problem was addressed using posterior sampling *via* Thompson-Sampling, identified as performant in balancing exploration and exploitation (Russo and Van Roy, 2014; Russo et al., 2018). The definition of the state was based on several individual's features including contextual information or sensor data from wearable devices while the reward was defined as the step counts within 30 min after the "dosing" event. The middle column of Table 1 summarizes the main characteristic of RL application to this problem.

3 Reinforcement learning in computational psychiatry

Like mechanistic modelling, computational psychiatry refers to a systems approach aimed at integrating underlying pathophysiological processes. However, while mechanistic modelling efforts typically use multiscale biological processes as building blocks, some models that fall within the remit of computational psychiatry (such as RL) use different types of building blocks, and in particular brain cognitive processes.

Model-based approaches have shown relevance for addressing major challenges in neuroscience (see (Conrado et al., 2020) for an example for Alzheimer disease). Quantitative systems pharmacology and mechanistic-based multiscale modelling are, in particular, associated with major hopes while acknowledging significant challenges such as the lack of quantitative and validated biomarkers, the subjective nature of clinical endpoints and the high selectivity of drug candidates not reflecting the complex interactions of different brain circuits (Geerts et al., 2020; Bloomingdale et al., 2021). These challenges are equally valid for attempting to address psychiatric conditions. This can partly explain the efficiency of non-pharmacological interventions, such as targeted psychotherapy approaches, recognized as one of the most precise and powerful approaches (Insel and Cuthbert, 2015).

The efficiency of such interventions is a testimony of how the brain's intrinsic plasticity can alter neural circuits. Some (discursive) disease models—with a focus on systems dimensions–propose new perspectives in the understanding of such conditions. For instance, it has been reported that emotion-cognition interactions gone awry can lead to anxiety and depression conditions; with anxious individuals displaying attentional-bias toward threatening stimuli and have difficulty disengaging from it (Crocker et al., 2013). Further data-driven understanding—at the systems level—is key to increase the likelihood of success of such non-pharmacological interventions, as it is equally the case for research and development of pharmaceutical compounds (Pao and Nagel,

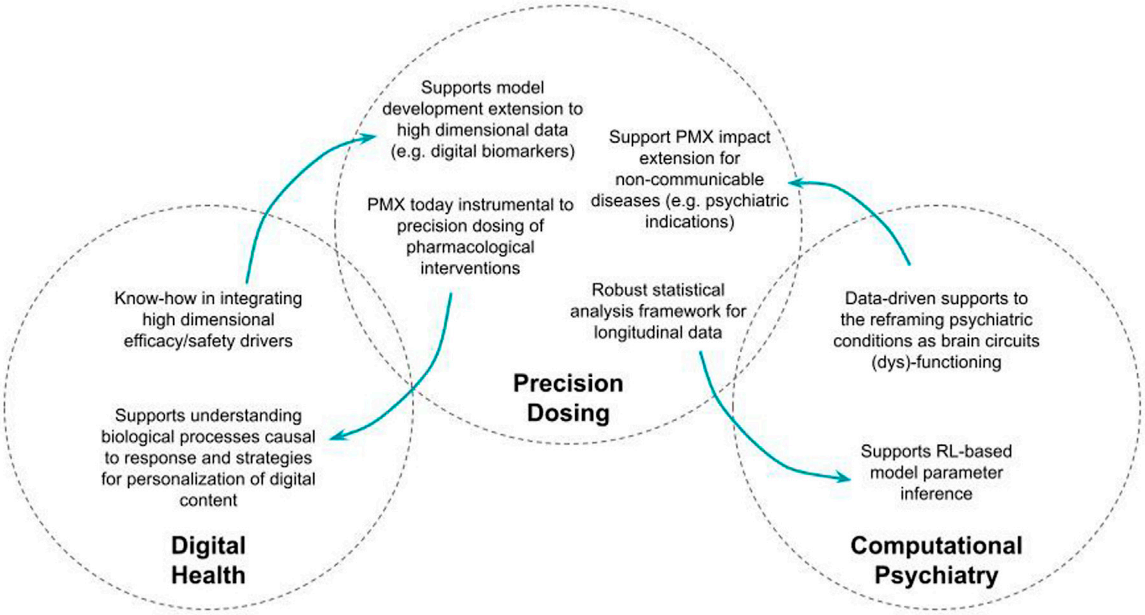
Supports model development extension to high dimensional data (e.g. digital biomarkers)
Support PMX impact extension for non-communicable diseases (e.g. psychiatric indications)
PMX today instrumental to precision dosing of pharmacological interventions
Know-how in integrating high dimensional efficacy/safety drivers
Robust statistical analysis framework for longitudinal data
Data-driven supports to the reframing psychiatric conditions as brain circuits (dys)-functioning
Precision Dosing
Supports understanding biological processes causal to response and strategies for personalization of digital content
Supports RL-based model parameter inference
Digital Health
Computational Psychiatry

FIGURE 1
Illustration of the mutual benefits of increased permeability between model-based approaches to precision dosing and digital health, on one hand, and computational psychiatry on the other hand.

2022). Such data-driven understanding can be integrated in the design of relevant non-pharmacological interventions, with some of them known to be amenable to digital delivery through, for instance, digital therapeutics (Jacobson et al., 2022).

A precision medicine initiative—precision psychiatry—has been initiated for psychiatric indications, such as major depression or substance abuse disorder, constituting a major part of non-communicable diseases (Insel and Cuthbert, 2015). The core idea of precision psychiatry lies in the reframing the diagnosis and care of affected subjects by moving away from a symptom-based to a data-driven categorization through a focus on system dimension *via* integration of data from cognitive, affective and social neuroscience, overall shifting the way to characterize these conditions in terms of brain circuits (dys-)functioning. This concept materialized in proposing the Research Domain Criteria (RDoc) in 2010 (Insel et al., 2010) as a framework for research in pathophysiology of psychiatric conditions.

Integrating into a multiscale modelling framework, data from cognitive, affective and social neuroscience is an objective of computational psychiatry, defined as a way to characterize mental dysfunction in terms of aberrant computation in the brain (Montague et al., 2012). Not surprisingly, by its mimicking of human and animal learning processes, RL plays a key role in computational psychiatry. RL in computational psychiatry proposes to map brain functioning in an algorithmic language offering then the possibility to explore, through simulations, the dysfunctioning of these processes as well as the theoretical benefit of interventional strategies. Two examples will be further developed here and the readers can refer to (Seriès, 2020) for an overview of more computational psychiatry methods, models and study cases.

In a RL framework, actions by the learner are chosen according to their value function, which holds the expected accumulated reward. The value function is updated through experience using feedback from the environment to the action taken. This update is also called temporal difference. An analogy has been drawn between this temporal difference and reward-error signals carried by dopamine in decision-making. Temporal difference reinforcement learning algorithms learn by estimating a value function based on temporal differences. The learning stops as this different converges to zero (see Supplementary Material for further details). Such a framework can be used to reframe addiction as a decision-making process gone awry. Based on the observation that addictive drugs produce a transient increase in dopamine through neuropharmacological mechanisms, the proposed model assumes that an addictive drug produces a positive temporal difference independent of the value function so that the action of taking drug will be always preferred over other actions (Redish, 2004). This model provides a tool to explore the efficiency of public health strategies. For instance, the model proposes some hypotheses to explain the incomplete success of strategies based on offering money as an alternate choice from drug intake.

RL models are used for the analysis of data of cognitive tasks and in particular tasks related to decision-making. Instead of focusing on the summary statistics of such tests (e.g, total number of errors), RL-based approaches allow for the integration of trial-by-trial data similarly to what model-based approaches typically do—with longitudinal data analysis—to better decipher response variability *via* the characterization of PK and PD processes. In the same way, trial-by-trial data can be leveraged to estimate RL-model based parameters which, in turn, can be compared to clinical endpoints such as measures of symptom severity to disentangle the role of brain circuit mechanisms overall contributing to a better understanding of response variability. RL for cognitive testing data in psychiatric populations is a complete paradigm change with respect to its application for precision dosing problems. While–in the two previous examples—RL was used to solve the problem of optimal dosing, now the RL algorithm is mapped to neuro-cognitive processes. Quantitatively characterizing these processes for each patient (estimating parameters from RL algorithms) is proposed as a methodology for extracting relevant information towards disease characterization and thus, response variability.

In (Huys et al., 2013), the authors use RL models to analyse population data of a behavioural test (signal-detection task) to study aspects of anhedonia—a core symptom of depression—related to reward learning. The authors proposed a RL model based on Q-learning update integrating two parameters: the classical learning rate and a parameter related to reward sensitivity modulating the percentage of the reward value actually contributes to the update of the Q value function. By performing a correlation analysis of the inferred parameters with anhedonic depression questionnaire, the authors found a negative correlation between the reward sensitivity but no correlation with the learning rate. Overall, these results led to the conclusion that the sensitivity to the reward and not the learning rate could be the main driver explaining why in anhedonic individuals, reward has less impact than in non-anhedonic individuals. Unravelling these two mechanisms is important for the planning of successful digital, behavioural and pharmacological strategies. The right column in Table 1 depicts the summary characteristics of RL applied to that study.

4 Conclusion

In this perspective, we have illustrated the flexibility of RL framework throughout the described applications in precision dosing, digital health and computational psychiatry and with that have demonstrated the benefit for the modeling community to become familiar with these approaches. The contrary is also true, and the field of precision digital therapeutics and computational psychiatry can benefit much from a proximity to the PMX community.

First, PMX methods could make RL even better. The field of computational psychiatry could benefit from input from the PMX community when it comes to statistical aspects related to parameters inference and clinical endpoint modelling. Two areas for which PMX has adopted as its state-of-the-art, population approach (with powerful algorithms such as stochastic approximation expectation-maximization algorithm (Lavielle, 2014)) and joint modelling respectively.

Second, the field of digital health should benefit from what constitutes one of the essential objectives of model-based drug development approaches, namely: elucidating response variability. It is particularly important for the successful development of digital therapeutic interventions to know how to characterize the efficacy and safety profiles and to know how to develop personalization strategies based on this understanding. The fact that it is about digital interventions should not prevent developers from prioritizing research in understanding underlying causal biological and (patho)-physiological processes of response, which will always be a key factor of successful therapy development, either pharmacological or not. Figure 1 proposes an illustration of these mutual benefits.

5 Legend

Table 1: Main characteristics of RL algorithm implementation to the precision dosing of pharmacological interventions (left column); the precision dosing of digital intervention (middle column); and computational psychiatry (right column). While there are multiple similarities between the precision dosing of pharmacological and digital interventions, the application of RL in computational psychiatry shows as a paradigm shift. RL computational machinery is not deployed as a technical approach to address the optimal control problem of precision dosing but is fitted to (cognitive task) data assuming the algorithm itself present mechanistic similarities with how brain's participants functioned during the task.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

BR: manuscript writing.

Acknowledgments

The author wishes to acknowledge Lucy Hutchinson, Richard Peck and Denis Engelmann for providing inputs on the drat manuscript.

Conflict of interest

The author is employed by F. Hoffmann La Roche Ltd.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.1094281/full#supplementary-material

References

Bloomingdale, P., Karelina, T., Cirit, M., Muldoon, S. F., Baker, J., McCarty, W. J., et al. (2021). Quantitative systems pharmacology in neuroscience: Novel methodologies and technologies. *CPT Pharmacometrics Syst. Pharmacol.* 10 (5), 412–419. doi:10.1002/psp4.12607

Bram, D. S., Parrott, N., Hutchinson, L., and Steiert, B. (2022). Introduction of an artificial neural network-based method for concentration-time predictions. *CPT Pharmacometrics Syst. Pharmacol.* 11 (6), 745–754. doi:10.1002/psp4.12786

Claes, S. (2022), Mobile health revolution in healthcare: Are we ready? Metaforum position paper 17 2019 [cited 2022 October 10]; Available at: https://www.kuleuven.be/metaforum/visie-en-debatteksten/2019-mobile-health-revolution-in-healthcare.

Conrado, D. J., Duvvuri, S., Geerts, H., Burton, J., Biesdorf, C., Ahamadi, M., et al. (2020). Challenges in alzheimer's disease drug discovery and development: The role of modeling, simulation, and open data. *Clin. Pharmacol. Ther.* 107 (4), 796–805. doi:10.1002/cpt.1782

Crocker, L. D., Heller, W., Warren, S. L., O'Hare, A. J., Infantolino, Z. P., and Miller, G. A. (2013). Relationships among cognition, emotion, and motivation: Implications for intervention and neuroplasticity in psychopathology. *Front. Hum. Neurosci.* 7, 261. doi:10.3389/fnhum.2013.00261

Fleisch, E., Franz, C., and Herrmann, A. (2021). *The digital pill*.

Friberg, L. E., Henningsson, A., Maas, H., Nguyen, L., and Karlsson, M. O. (2002). Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J. Clin. Oncol.* 20 (24), 4713–4721. doi:10.1200/JCO.2002.02.140

Geerts, H., Wikswo, J., van der Graaf, P. H., Bai, J. P. F., Gaiteri, C., Bennett, D., et al. (2020). Quantitative systems pharmacology for neuroscience drug discovery and development: Current status, opportunities, and challenges. *CPT Pharmacometrics Syst. Pharmacol.* 9 (1), 5–20. doi:10.1002/psp4.12478

Henin, E., Meille, C., Barbolosi, D., You, B., Guitton, J., Iliadis, A., et al. (2016). Revisiting dosing regimen using PK/PD modeling: The MODEL1 phase I/II trial of docetaxel plus epirubicin in metastatic breast cancer patients. *Breast Cancer Res. Treat.* 156 (2), 331–341. doi:10.1007/s10549-016-3760-9

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., and Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biol. Mood Anxiety Disord.* 3 (1), 12. doi:10.1186/2045-5380-3-12

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 167 (7), 748–751. doi:10.1176/appi.ajp.2010.09091379

Insel, T. R., and Cuthbert, B. N. (2015). Medicine. Brain disorders? Precisely. *Science* 348 (6234), 499–500. doi:10.1126/science.aab2358

Jacobson, N. C., Kowatsch, T., and Marsch, L. A. (2022). *Digital therapeutics for mental health and addiction: The state of the science and vision for the future*. San Diego, CA: Academic Press, 270.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., et al. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychol.* 34S, 1220–1228. doi:10.1037/hea0000305

Lavielle, M. (2014). *Mixed effects models for the population approach: Models, tasks, methods and tools*. 1st edition. Chapman and Hall/CRC.

Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., et al. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *Lancet* 380 (9838), 219–229. doi:10.1016/S0140-6736(12)61031-9

Liao, P., Greenewald, K., Klasnja, P., and Murphy, S. (2020). Personalized HeartSteps: A reinforcement learning algorithm for optimizing physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4 (1), 18. doi:10.1145/3381007

Lu, J., Deng, K., Zhang, X., Liu, G., and Guan, Y. (2021). Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience* 24 (7), 102804. doi:10.1016/j.isci.2021.102804

Maier, C., Hartung, N., Kloft, C., Huisinga, W., and de Wiljes, J. (2021). Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology. *CPT Pharmacometrics Syst. Pharmacol.* 10 (3), 241–254. doi:10.1002/psp4.12588

Maxfield, K., and Zineh, I. (2021). Precision dosing: A clinical and public health imperative. *JAMA* 325 (15), 1505–1506. doi:10.1001/jama.2021.1004

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends Cogn. Sci.* 16 (1), 72–80. doi:10.1016/j.tics.2011.11.018

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., et al. (2018). Just-in-Time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* 52 (6), 446–462. doi:10.1007/s12160-016-9830-8

Pao, W., and Nagel, Y. A. (2022). Paradigms for the development of transformative medicines-lessons from the EGFR story. *Ann. Oncol.* 33 (5), 556–560. doi:10.1016/j.annonc.2022.02.005

Peck, R. W. (2021). Precision dosing: An industry perspective. *Clin. Pharmacol. Ther.* 109 (1), 47–50.

Qian, T., Klasnja, P., and Murphy, S. A. (2020). Linear mixed models with endogenous covariates: Modeling sequential treatment effects with application to a mobile health study. *Stat. Sci.* 35 (3), 375–390. doi:10.1214/19-sts720

Qian, T., Walton, A. E., Collins, L. M., Klasnja, P., Lanza, S. T., Nahum-Shani, I., et al. (2022). The microrandomized trial for developing digital interventions: Experimental design and data analysis considerations. *Psychol. Methods* 27, 874–894. doi:10.1037/met0000283

Redish, A. D. (2004). Addiction as a computational process gone awry. *Science* 306 (5703), 1944–1947. doi:10.1126/science.1102384

Ribba, B., et al. (2022). *Model enhanced reinforcement learning to enable precision dosing: A theoretical case study with dosing of propofol*. CPT Pharmacometrics Syst Pharmacol.

Ribba, B., Kaloshi, G., Peyre, M., Ricard, D., Calvez, V., Tod, M., et al. (2012). A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clin. Cancer Res.* 18 (18), 5071–5080. doi:10.1158/1078-0432.CCR-12-0084

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z. (2018). A tutorial on Thompson sampling. *Found. Trends® Mach. Learn.* 11 (1), 1–96. doi:10.1561/2200000070

Russo, D., and Van Roy, B. (2014). Learning to optimize via posterior sampling. *Math. Operations Res.* 39 (4), 1221–1243. doi:10.1287/moor.2014.0650

Seriès, P. E. (2020). *Computational psychiatry*. The MIT Press.

Sutton, R., and Barto, A. (2018). *Reinforcement learning: An introduction*. Second edition.

Sverdlov, O., van Dam, J., Hannesdottir, K., and Thornton-Wells, T. (2018). Digital therapeutics: An integral component of digital innovation in drug development. *Clin. Pharmacol. Ther.* 104 (1), 72–80. doi:10.1002/cpt.1036

Wang, D., Song, Z., Zhang, C., and Chen, P. (2021). Bispectral index monitoring of the clinical effects of propofol closed-loop target-controlled infusion: Systematic review and meta-analysis of randomized controlled trials. *Med. Baltim.* 100 (4), e23930. doi:10.1097/MD.0000000000023930

Yauney, G., and Shah, P. (2018). "Reinforcement learning with action-derived rewards for chemotherapy and clinical trial dosing regimen selection," in Proceedings of the 3rd Machine Learning for Healthcare Conference (PMLR: Proceedings of Machine Learning Research), 161–226. D.-V. Finale, et al., Editors.