FROM NOISE TO SIGNAL: UNVEILING TREATMENT EFFECTS FROM DIGITAL HEALTH DATA THROUGH PHARMACOLOGY-INFORMED NEURAL-SDE

Samira Pakravan[1, 2, +], Nikolaos Evangelou[1, 3, +], Maxime Usdin[4], Logan Brooks[1, *], and James Lu[1, *]

[1]Clinical Pharmacology, Genentech, South San Francisco, CA 94080, USA
[2]Department of Mechanical Engineering, University of California, Santa Barbara, CA, USA
[3]Department of Chemical and Biomolecular Engineering Johns Hopkins University Baltimore, MD, USA
[4]Computational Sciences, Genentech, South San Francisco, CA 94080, USA
[+]Equal contribution
[*]Corresponding authors (brooksl2@gene.com, lu.james@gene.com)

ABSTRACT

Digital health technologies (DHT), such as wearable devices, provide personalized, continuous, and real-time monitoring of patient. These technologies are contributing to the development of novel therapies and personalized medicine. Gaining insight from these technologies requires appropriate modeling techniques to capture clinically-relevant changes in disease state. The data generated from these devices is characterized by being stochastic in nature, may have missing elements, and exhibits considerable inter-individual variability - thereby making it difficult to analyze using traditional longitudinal modeling techniques. We present a novel pharmacology-informed neural stochastic differential equation (SDE) model capable of addressing these challenges. Using synthetic data, we demonstrate that our approach is effective in identifying treatment effects and learning causal relationships from stochastic data, thereby enabling counterfactual simulation.

1 INTRODUCTION

The rise of digital health technologies (DHT) including wearable devices such as smart watch and patch based physiological sensors has opened new possibilities for continuous patient monitoring Friend et al. (2023) and enables generation of time-series data at an unprecedented temporal resolution and duration, thereby offering the potential to generate new clinical measures and insights Berisha et al. (2021). Furthermore, recent examples have shown the clinical value in modeling both the longitudinal trends as well as the stochastistity in digital health (DH) data Leander et al. (2022).

Stochastic differential equations (SDEs) have been developed to describe various phenomena that exhibits random fluctuations Fagin et al. (2023), including in biological and biomedical applications Mei et al. (2013); Tajmirriahi & Amini (2021). In the context of DH, the interplay between physiology and the measurement device is likely far too complex for one to theoretically derive the equations underlying the link between disease status and DH data from first principles. Instead, we propose to learn the underlying dynamical system directly from data, with the help of neural-SDE Evangelou et al. (2023); Dietrich et al. (2023).

Here, we develop a *pharmacology-informed* Lu et al. (2021); Laurie & Lu (2023) neural-SDE that:

- learns the underlying dynamical system from a patient population, while introducing patient-dependent parameters that enables the characterization of patient-to-patient variability;
- incorporates the causality between pharmacokinetics (PK) and pharmacodynamics (PD);
- enables counterfactual simulations to describe drug effects at the individual patient level.

We demonstrate the effectiveness of the proposed model using synthetic data.

2 METHODS

2.1 PHARMACOLOGY-INFORMED NEURAL-SDE (PI-NSDE) MODEL

We assume that the longitudinal data are modelled by a system of equations of the form,

$$dc_t = f(c_t)dt \tag{1}$$
$$dx_t = \nu(x_t, c_t, \boldsymbol{p})dt + \sigma(x_t, c_t, \boldsymbol{p})dW_t \tag{2}$$

where Equation 1 represents a known Ordinary Differential Equation (ODE) model with $f(\cdot)$ being the vector field for PK that governs the drug concentration, $c_t \in \mathbb{R}$, and where the drift and diffusion terms (i.e., $\nu(x_t, c_t, \boldsymbol{p})$ and $\sigma(x_t, c_t, \boldsymbol{p})$ respectively) are described by neural networks.

We work under the hypothesis that the drift and diffusivity terms of the *effective* SDE, are dependent on the state ($x_t \in \mathbb{R}$) as well as the drug concentration c_t. Additionally, while the underlying equations are the same for all patients, the model includes a *latent* patient-dependent parameter vector $\boldsymbol{p}$ that describes the patient-to-patient variability. This *latent* parameter $\boldsymbol{p}$ is discovered in a data-driven way based on the work of Lu et al. (2021), which we elaborate below.

While the available data are in the form of trajectories, we transform them to snapshots $\mathcal{D}$ in a manner analogous to that done in Dietrich et al. (2023). In particular, each snapshot $\mathcal{D}^i$, uniquely identified by the index i, takes the form $\mathcal{D}^i = \{x_1^i, x_0^i, \Delta t, c_1^i, \boldsymbol{p}^{i,j}\}$, where x_1^i is the evolution of the state variable x_t after a time step Δt given the initial condition x_0^i; $\boldsymbol{p}^{i,j}$ is the *latent* parameter for the jth patient. Note that we utilize the concentration at c_1 and not at c_0 following the (symplectic) Euler-Maryama scheme discussed in Dietrich et al. (2023). The concentration c_t and the patient dependent parameter $\boldsymbol{p}$ enter into the overall architecture as inputs based on Dietrich et al. (2023); Evangelou et al. (2023).

The construction of the loss function (based on Dietrich et al. (2023)) is derived from the numerical integration scheme (symplectic) Euler-Maruyama. The numerical approximation of Equations 1 and 2 results in:

$$c_1^i = c_0^i + f(c_0^i)\Delta t \tag{3}$$
$$x_1^i = x_0^i + \nu(x_0^i, c_1^i, \boldsymbol{p}^{i,j})\Delta t + \sigma(x_0^i, c_1^i, \boldsymbol{p}^{i,j})\delta W_0, \tag{4}$$

where δW_0 is normally distributed around zero and Δt is a variable timestep. The drift and diffusivity terms are approximated by two networks ν_θ and σ_θ, under the assumption that x_1 is drawn from a normal distribution of the form,

$$x_1^i \sim \mathcal{N}(x_0^i + \nu_\theta(x_0^i, c_1^i, \boldsymbol{p}^{i,j})\Delta t, \sigma_\theta(x_0^i, c_1^i, \boldsymbol{p}^{i,j})^2 \Delta t). \tag{5}$$

With the assumed mean and variance in Equation 5 for the drift and diffusivity, we can compute the logarithm of the resulting normal distribution and derive the following loss function that maximizes the likelihood:

$$\mathcal{L}(\theta|x_0^i, x_1^i, \Delta t) := \frac{(x_1^i - x_0^i - \nu_\theta(x_0^i, c_1^i, \boldsymbol{p}^{i,j}))^2}{\Delta t \sigma_\theta(x_0^i, c_1^i, \boldsymbol{p}^{i,j})^2} + \log|\Delta t \sigma(x_0^i, c_1^i, \boldsymbol{p}^{i,j})^2|. \tag{6}$$

It should be noted that the Neural-SDE framework by Dietrich et al. (2023) is also capable of handling varying time steps Δt.

The Neural-SDE architecture consists of two network components for the drift and diffusion models. In our work, the drift network consists of 4 layers where each layer has 64 neurons each followed by `ELU` activation function. The diffusion network consists of 3 layers with 32 neurons, the first two layers are followed by `ELU` activation function and the output layer is followed by `softplus` activation function. A schematic of the Neural-SDE architecture is shown in Figure 1.

2.2 Latent Patient Descriptors - GRU Encoder

Our approach to learning the Neural-SDE from data across the patient population is to identify a set of dynamical equations that holds across all patients, as well as patient-specific descriptors (or embedding) that characterize patient-to-patient variability Laurie & Lu (2023). In our approach, those patient-specific descriptors are discovered in a data-driven manner, based on the work of Lu et al. (2021): a Gated Recurrent Unit (GRU) encoder was used to discover the latent parameter p, with longitudinal data provided in a tabular form as an input. More specifically, the input data entering the encoder consist of variable number of rows for each patient and the following four columns: (1) the absolute time; (2) the time after dose; (3) the stochastic PD data (4) the deterministic PK data.

Each tabular input was *padded* and *masking* was applied in order to handle the variable time points. The GRU encoder has 128 hidden states and is connected to a Multilayer Perceptron (MLP) consisting of 2 layers, each with 128 neurons, both followed by `ELU` activation function. The output of MLP is the *latent* parameter p that enters the Neural-SDE architecture. An end-to-end training was implemented by using the loss function given by Equation 6.
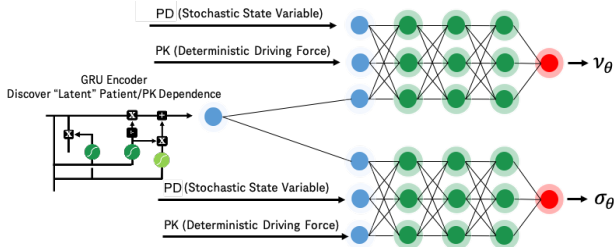

Figure 1: The Neural-SDE architecture including the GRU encoder.

2.3 Dataset

To mimic clinical digital health measurements, synthetic data was simulated in which the PK serves as a deterministic driving input that causally influences a stochastically evolving PD. Patient specific parameters were sampled from a log-normal distribution: 50 individual patient trajectories were sampled across 3 different dose levels (50 mg, 100 mg, 400 mg) for a total of 150 patient trajectories and 70:30 train-test split was used; further details are summarized in Appendix A.1.

3 Results

Table 1 presents the summary of results and a comparison of model performance with and without inclusion of the patient specific representation learned by the GRU. Overall, model performance is high with improved model performance with the GRU, notably, when looking at the model's ability to capture the population and individual dose / exposure response relationships. Figure 2 demonstrates the model's ability to learn the underlying system's dynamics by comparing "true" (i.e., the underlying ground truth) SDE trajectories from the test dataset against the model predicted trajectories. For each patient in the test set, we sampled 250 trajectories to provide a robust representation of the predictive variability associated with the model. This result demonstrates the model's ability in replicating the complex dynamics of PD trajectories at the population level.

Figure 2: Comparison of the true and predicted SDE trajectories in the test dataset. Left panel: the colored lines represent the observed stochastic trajectories in the test data. Right panel: blue line and shaded region represent the median and the 10th to 90th percentile respectively of the ground truth trajectories; similarly, the orange lines and shade region represent those from the model.

3.1 DOSING REGIMEN ANALYSIS

To analyze the impact of different dosing regimens on PD, we consider three distinct simulated doses at $50\,\mathrm{mg}$, $100\,\mathrm{mg}$, and $400\,\mathrm{mg}$. For each patient from the test dataset, we sampled 250 SDE trajectories. Figure 3 shows the the model is qualitatively able to capture the true underlying dose response relationship.
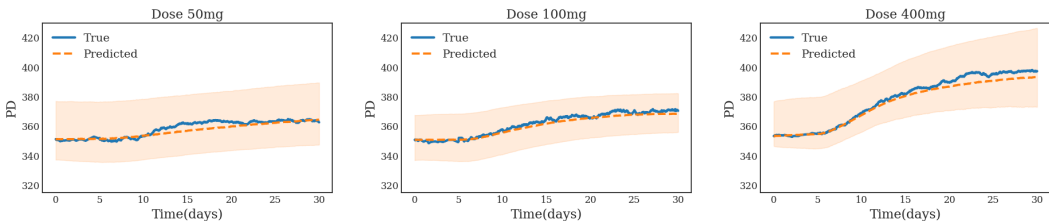

Figure 3: Comparison of the true and predicted SDE trajectories in the test datase for 50, 100 and 400 mg doses. Blue lines represent the median of the ground truth trajectories; orange dashed lines and shaded regions represent median and the 10th to 90th percentile of trajectories from the model.

3.2 PATIENT-SPECIFIC RESPONSES AND COUNTERFACTUAL ANALYSIS

Figure 4 demonstrates the proposed methodology's ability to perform counterfactual analysis and identify individual treatment effects. To accomplish this, for each patient the drift and diffusivity terms were inferred from the trained model and 250 SDE trajectories were generated. The results demonstrate the model's ability to capture the underlying dynamics of the stochastic process for individual patients. This suggests that the GRU encoding strategy not only captures the population behaviors, but also successfully learns to differentiate amongst patients. Moreover, we demonstrate a what-if scenario: in the absence of PK, the model correctly predicts a lack of dynamical change in the modeled PD endpoint. This suggests our model is able to correctly identify the causal relationship between PK and PD.
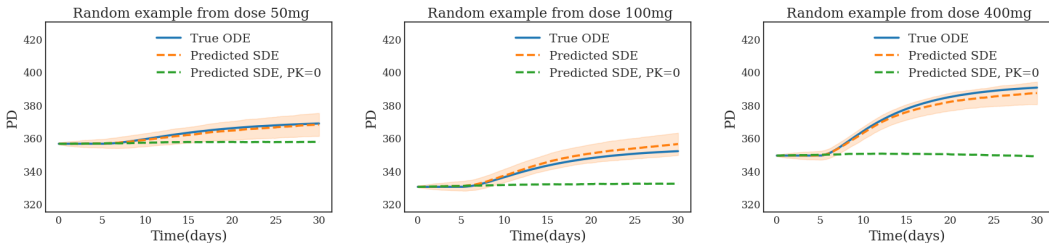
Figure 4: Patient-specific trajectories and counterfactual simulations. Each subplot represents a random patient from the respective dosages. The solid blue line represents the true drift; the orange dashed line and shaded region represent the mean and mean $\pm$ standard deviation (std) of 250 posterior samples; the green dashed lines represent counterfactual simulations assuming no dosing (*i.e.*, PK = 0).

Table 1: Summary of results and comparison between PI-NSDE model with the patient specific representation (+GRU) and without the patient specific representation (-GRU).

Model	Model Variant	Dose Level / Experiment	MSE	R^2
PI-NSDE	(-GRU)	Overall	16.03	0.95
	(+GRU)		**5.9**	**0.98**
PI-NSDE	(-GRU)	Population: 50mg	3.23	0.85
	(+GRU)		**1.22**	**0.94**
PI-NSDE	(-GRU)	Population: 100mg	10.82	0.82
	(+GRU)		**1.50**	**0.98**
PI-NSDE	(-GRU)	Population: 400mg	17.51	0.93
	(+GRU)		**2.65**	**0.99**
PI-NSDE	(-GRU)	Individual: Dose=50	5.45	0.97
	(+GRU)		**1.79**	**0.99**
PI-NSDE	(-GRU)	Individual: Dose=100	9.83	0.92
	(+GRU)		**8.68**	**0.93**
PI-NSDE	(-GRU)	Individual: Dose=400	40.03	0.86
	(+GRU)		**9.40**	**0.97**
PI-NSDE	(-GRU)	Counter Factual	1.96	0.99
	(+GRU)		1.92	0.99

4 CONCLUSION

We proposed a pharmacology-informed neural-SDE architecture that is able learn the relationship between a deterministic PK and stochastic PD. Using synthetic data, the model correctly reproduces the underlying PK-PD relationship at the population level. Furthermore, the model enables the counterfactual simulation of PD in the absence of the hypothetical drug - and in doing so, quantify the individual treatment effect. Future work will evaluate this modeling approach on wearable ECG data collected in clinical trial patients as well as applying it in other clinical settings where consideration for the stochasticity of the data could provide additional insights.

REFERENCES

Visar Berisha, Chelsea Krantsevich, P Richard Hahn, Shira Hahn, Gautam Dasarathy, Pavan Turaga, and Julie Liss. Digital medicine and the curse of dimensionality. *NPJ digital medicine*, 4(1):153, 2021.

N. L. Dayneka, V. Garg, and W. J. Jusko. Comparison of four basic models of indirect pharmacodynamic responses. *Journal of pharmacokinetics and biopharmaceutics*, 21(4):457–478, 1993. doi: 10.1007/BF01061691.

Felix Dietrich, Alexei Makeev, George Kevrekidis, Nikolaos Evangelou, Tom Bertalan, Sebastian Reich, and Ioannis G Kevrekidis. Learning effective stochastic differential equations from microscopic simulations: Linking stochastic numerics to deep learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(2), 2023.

Nikolaos Evangelou, Felix Dietrich, Juan M Bello-Rivas, Alex J Yeh, Rachel S Hendley, Michael A Bevan, and Ioannis G Kevrekidis. Learning effective sdes from brownian dynamic simulations of colloidal particles. *Molecular Systems Design & Engineering*, 2023.

Joshua Fagin, Ji Won Park, Henry Best, KE Saavik Ford, Matthew J Graham, V Ashley Villar, Shirley Ho, James Hung-Hsu Chan, and Matthew O'Dowd. Latent stochastic differential equations for modeling quasar variability and inferring black hole properties. *arXiv preprint arXiv:2304.04277*, 2023.

Stephen H Friend, Geoffrey S Ginsburg, and Rosalind W Picard. Wearable digital health technology, 2023.

Mark Laurie and James Lu. Explainable deep learning for tumor dynamic modeling and overall survival prediction using neural-ode. *npj Systems Biology and Applications*, 9(1):58, 2023.

Jacob Leander, Mats Jirstrand, Ulf G Eriksson, and Robert Palmér. A stochastic mixed effects model to assess treatment effects and fluctuations in home-measured peak expiratory flow and the association with exacerbation risk in asthma. *CPT: Pharmacometrics & Systems Pharmacology*, 11(2):212–224, 2022.

James Lu, Brendan Bender, Jin Y Jin, and Yuanfang Guan. Deep learning prediction of patient response time course from early data via neural-pharmacokinetic/pharmacodynamic modelling. *Nature machine intelligence*, 3:696–704, 2021.

Yongguo Mei, Adria Carbo, Raquel Hontecillas, and Josep Bassaganya-Riera. Enisi sde: a novel web-based stochastic modeling tool for computational biology. In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 392–397. IEEE, 2013.

Mahnoosh Tajmirriahi and Zahra Amini. Modeling of seizure and seizure-free eeg signals based on stochastic differential equations. *Chaos, Solitons & Fractals*, 150:111104, 2021.

A   APPENDIX

A.1   DATASET GENERATION DETAILS

Synthetic training data was generated to represent wearable ECG data which is being used to evaluate a drug candidate's potential to prolong the QT interval. This evaluation is done for all clinical drug candidates as QT prolongation increases the risk of a drug to cause fatal arrhythmia such as Torsades de Pointes. This synthetic data was characterized by indirect response PK-PD model Dayneka et al. (1993) by which PK acts causally to change the PD, with the additional modification that the observable PD variable is stochastic in nature. This system follows the general form of Equations 1 and 2, with the following system of ODEs being specified for the term $f(c_t, \boldsymbol{p})dt$:

$$\frac{du_1}{dt} = -KA \times u_1(t) \tag{7}$$

$$\frac{du_2}{dt} = KA \times u_1(t) - u_2(t) \times (KE + K12) + u_3(t) \times K21 \tag{8}$$

$$\frac{du_3}{dt} = K12 \times u_2(t) - K21 \times u_3(t) \tag{9}$$

where $c_t = u_2(t)/\text{V2}$ with V2 representing the volume of distribution for drug in plasma circulation. The drift term in the relationship between c_t and PD is represented by the following:

$$\frac{du_4}{dt} = KIN - (KOUT * (1 - (Imax \times c_t/IC50 + c_t))) \times u_4(t). \tag{10}$$

Example trajectories of this system are shown in Figure 5. The diffusion term in Equation 2 is described by the following $\beta u_4 dW_t$, where β was sampled from a log-normal distribution.

Examples of stochastic trajectories for c_t are shown in Figure 6.

In the current set of experiments, an equal number of patients were simulation for a range of doses (50, 100, 400 mg). Dosing was set to begin at day 5 for all synthetic subjects with daily dosing; the PD sampling frequency is once per hour, over a period of 30 days.
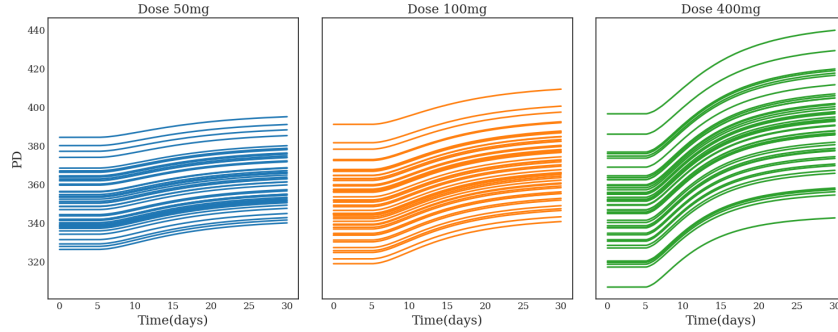

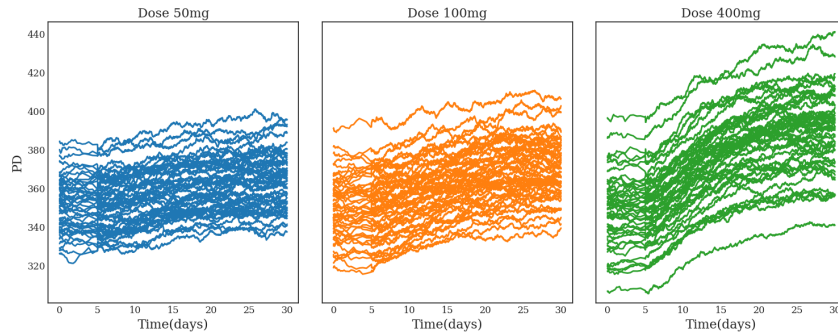Figure 5: Synthetic data trajectories without the diffusivity component under different simulated doses.

Figure 6: Synthetic data trajectories under different doses.

A.2 TRAINING METHODOLOGY AND OPTIMIZATION STRATEGY

The current model, including the numerical integration scheme which employs a Euler-Maruyama solver, have been implemented in PyTorch. While a higher-order methods were not used in this current work, it remains open for future development based on specific needs.

In model training, we leveraged vectorization rather than operating on a single value at a time whereby the model processes each time-step for each patient sequentially. In this way, the model operates at a patient level, concurrently processing all data points associated with a specific patient. This is feasible based on the observation that evaluating the loss function given in Equation 6 at each time-step is independent from other time instances. The vectorization strategy significantly enhances the training and inference performance.

We trained the network for 100 epochs using the ADAM optimizer with learning rate 0.001 and batch size of 1. The overall training process takes around 140 seconds using one NVIDIA V100 GPU.