

Announcement of Machine Learning Benchmarking

Mark Sale (1), Liang Zhao (2)

Institution: (1) Certara (2) University of California, San Francisco

Background & Objectives

Automated methods for population pharmacokinetic model selection has recently been an area of significant progress. Among the method that the authors have identified are:

Neural ordinary differential equations [1,2,3], a method for sequential selection of model components (e.g., structure, covariate, between subject variability) [4], unsupervised machine learning methods such as Genetic Algorithm [3,5], Random Forests and Gaussian Process [5], neural networks and support vector machines [6].

Methods

We propose a prospective, blinded study of the performance of available tools. A similar approach was undertaken to comparison parameter estimation algorithms and presented at PAGE in 2005 [7]. The parameter estimation benchmark was done only a single time. As the automated methods are currently rapid advancing, we propose a biannual benchmark exercise, similar to the The Critical Assessment of Structure Prediction (CASP) [8] of protein folding algorithms.

Participants and organizers will be invited from across the global pharmacometrics (PMX) community to ensure broad engagement and representation. An initial organizing team—potentially coordinated by UCSF—will draft preliminary ideas for the test problems, search space, and model evaluation criteria, with the intent of collaboratively refining these elements through community input.

Methods (continued)

The overarching goal is to define inclusive and widely applicable criteria that accommodate a diversity of modeling approaches and performance metrics.

While UCSF may help catalyze the process, the final design of the test problems, search space, and evaluation criteria will be shaped with input from multiple stakeholders. Evaluation metrics will be multi-dimensional, addressing various facets of model performance beyond a single outcome.

To further enhance transparency and trust, the assessment of submissions may be conducted by a neutral third party or an independent review committee composed of experts from different institutions. Tentatively, the first-year challenge problems—including data sets, defined search space, and evaluation framework—may be shared with participants around October 31, 2025, with a provisional deadline for results submission by January 2026. These timelines are subject to change depending on broader consensus and logistical readiness.

Final results are expected to be presented at a major PMX community event, such as the PAGE meeting in 2026.

[1] Lu, James et al. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience*, Volume 24, Issue 7, 102804

[2] Dominic Stefan Bräm, Uri Nahum, Johannes Schropp et al. Neural ODEs in Pharmacokinetics: Concepts and Applications, 06 January 2023, PREPRINT (Version 1) available at Research Square

[3] Population pharmacokinetic model selection assisted by machine learning. *J Pharmacokinet Pharmacodyn* 49, 257–270 (2022).

[4] A fully automatic tool for development of population pharmacokinetic models. *CPT Pharmacometrics Syst Pharmacol*. 2024 Oct;13(10):1784-1797.

[5] pyDarwin: A Machine Learning Enhanced Automated Nonlinear Mixed-Effect Model Selection Toolbox. *Clin Pharmacol Ther*, 2024,115: 758-773.

[6] Population Pharmacokinetic and Pharmacodynamic Models of Propofol in Healthy Volunteers using NONMEM and Machine Learning Methods. *Health Inform Res*. 2008;14(2):147-159. Published online June 30, 2008

[7] <https://www.page-meeting.org/?abstract=834> accessed 15 May 2025

[8] <https://predictioncenter.org/> accessed 15 May, 2025