# Machine learning did not outperform NLME-based Bayesian forecasting for early detection of delayed methotrexate elimination

**Marian Klose**[1,2], Dominik Marschner[3], Markus Knott[3,4], Julia Wendler[3,4], Julian Müller-Kühnle[3], Lukas Kovar[5], Christin Nyhoegen[1], Andreas Meltzow[1,2], Niklas Hartung[6], Wilhelm Huisinga[2,6], Robin Michelet[1], Anna M. Mc Laughlin[1,7], Gerald Illerhaus[3,4], Charlotte Kloft[1,2]

[1]Freie Universität Berlin, Institute of Pharmacy, Department of Clinical Pharmacy and Biochemistry, Berlin, Germany; [2]Graduate Research Training Program PharMetrX, Berlin/Potsdam, Germany; [3]Department of Haematology/Oncology and Palliative Care, Klinikum Stuttgart, Stuttgart, Germany; [4]Stuttgart Cancer Center – Tumorzentrum Eva Mayr-Stihl, Klinikum Stuttgart, Stuttgart, Germany; [5]Clinical Pharmacology and Safety Science, Boehringer Ingelheim Pharma GmbH & Co. KG; [6]Institute of Mathematics, University of Potsdam, Potsdam, Germany; [7]Pharmetheus AB, Uppsala, Sweden

## Background

**HD-MTX:** renally eliminated and nephrotoxic → risk of delayed elimination ($t_{0.2\mu M}$ > 72 h) → severe toxicities[1]

**Early detection:** crucial for timely rescue interventions[2]

**Bayes-NLME:** limited early predictivity (0h/4h; sparse data)[3]
**ML:** proposed alternative for improved early predictions[4]

## Objective

**To compare whether machine learning improves early prediction of delayed MTX elimination compared to NLME-based full Bayesian forecasting**

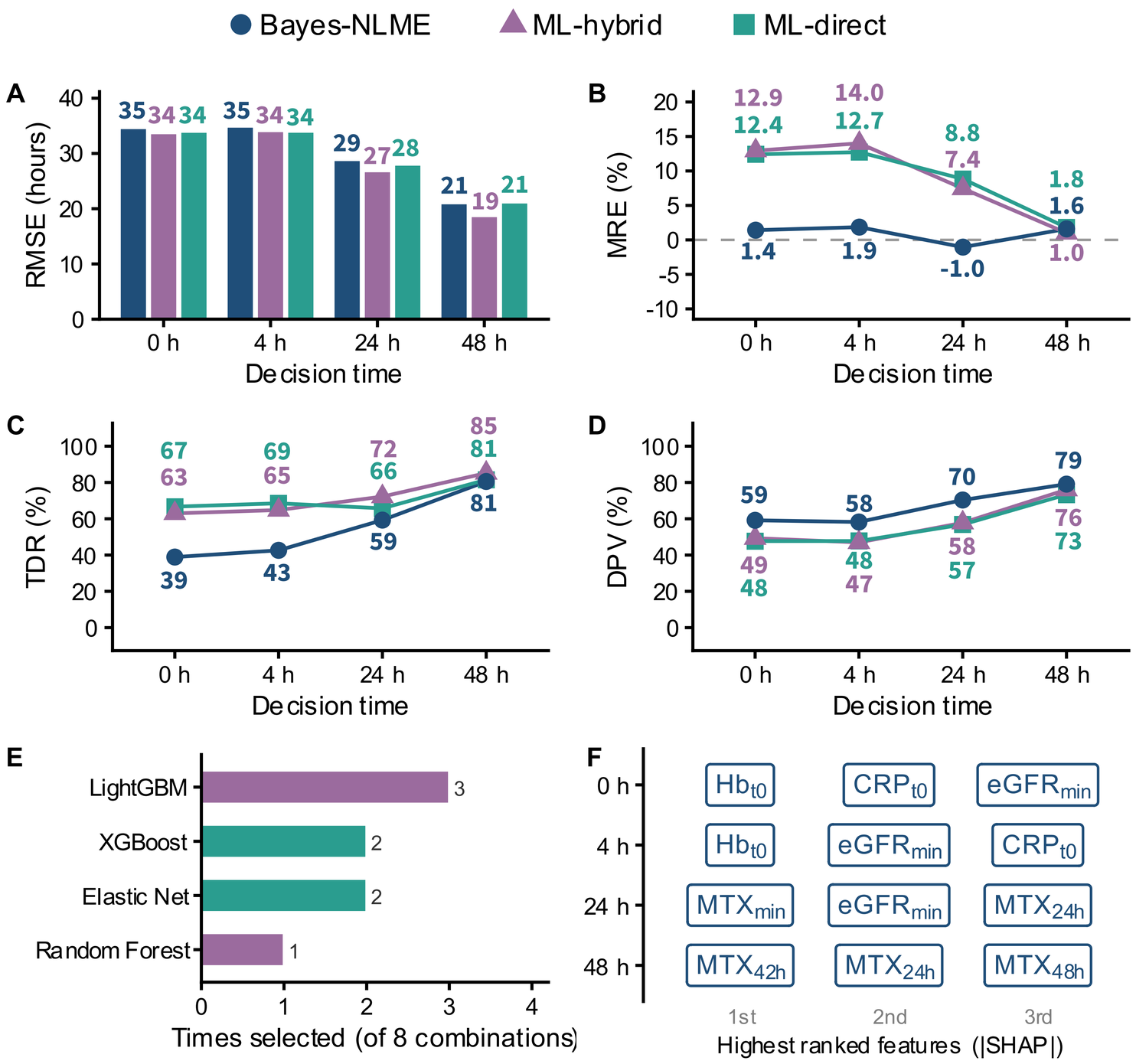
## Methods

1 ML algorithm selected per decision time × prediction mode combination (n = 8)

| Clinical routine data | Features | Prediction modes | ML algorithms | Tuning | Evaluation |
|---|---|---|---|---|---|
| 176 adult CNS-lymphoma patients | Demographics | Direct: $t_{0.2\mu M}$ prediction | XGBoost, LightGBM | 5×10-fold cross-validation | Continuous: RMSE, MRE |
| 547 HD-MTX cycles | Biomarkers, TDM | Hybrid: corrects Bayes-NLME $t_{0.2\mu M}$ | RF, Elastic net | Space-filling grid (n = 100, latin-hypercube design) | Categorical: TDR, DPV |
| 75/25 ID-based split | Prior-cycle history | Decision times: 0 h / 4 h / 24 h / 48 h | Cubist, MARS | Selection by RMSE | SHAP analysis |
| | Dosing information | | SVR, k-NN, ffNN | | |

*Framework: tidymodels*

## Results & Interpretation*




Comparable overall predictive performance (ΔRMSE ≤ 2.33 h)

Bayes-NLME unbiased; ML overpredicted $t_{0.2\mu M}$ at 0-24 h; Similar performance at 48 h

ML flagged more delayed eliminators (↑TDR) at the cost of lower reliability (↓DPV)

Gradient-boosted trees (LightGBM, XGBoost) selected most often (5/8)

Top |SHAP| features: covariates at 0/4 h, TDM samples at 24/48 h

**All panels show out-of-fold cross-validated predictions from the training cohort; the independent ID-based test cohort is not shown and is reserved for final validation.*

[1]Methotrexate Injection Prescribing Info. Accord Healthcare Inc.; 2022.
[2]Bielack et al. J Cancer Res Clin Oncol. 2024;150:441.
[3]Klose et al. PAGE 33 (2025), Abstr. 11507.
[4]Destere et al. Clin Pharmacokinet. 2022;61:1157–1165.

**Abbreviations**: CNS, central nervous system; CRP, C-reactive protein; CV, cross-validation; DPV, delayed predictive value; eGFR, estimated glomerular filtration rate; ffNN, feed-forward neural network; Hb, hemoglobin; HD-MTX, high-dose methotrexate; k-NN, k-nearest neighbors; LightGBM, light gradient-boosting machine; MARS, multivariate adaptive regression splines; ML, machine learning; MRE, mean relative error; NLME, nonlinear mixed-effects; RF, random forest; RMSE, root mean squared error; SHAP, SHapley Additive exPlanations; SVR, support vector regression; $t_{0.2\mu M}$, time until MTX concentration falls below 0.2 µM; TDR, true delayed rate; XGBoost, extreme gradient boosting.

*For more information:*
**Marian Klose**
marian.klose@fu-berlin.de
www.clinical-pharmacy.eu
www.marian-klose.com
website

*34th Population Approach Group Europe meeting – PAGE, Dubrovnik, Croatia 2026*