

Article

Evaluation and Benchmarking of a Bounded Data-Driven Correction for Compartmental Pharmacokinetic Models

Hanan Al Lawati ¹, Abdullah Al Lawati ² and Mohamed Al-Lawatia ^{3,*}

¹ Pharmacy Program, Oman College of Health Sciences, P.O. Box 3720, Ruwi 112, Oman; hanan.allawati@ochs.edu.om

² Department of Surgery, Sultan Qaboos University Hospital, P.O. Box 50, Al-Khodh 123, Oman; abdullah.allawati.00@gmail.com

³ Department of Mathematics, Sultan Qaboos University, P.O. Box 50, Al-Khodh 123, Oman

* Correspondence: allawati@squ.edu.om

Abstract

Background/Objectives: This study applies the previously introduced structure-preserving hybrid framework for compartmental pharmacokinetic models and extends its clinical evaluation using published clinical datasets. The framework combines a mechanistic pharmacokinetic backbone with a bounded data-driven correction. The aim was to assess predictive performance in held-out subjects while keeping the main pharmacokinetic structure and avoiding a fully black-box model. **Methods:** This applied extension of the framework was tested in several numerical studies using published clinical pharmacokinetic datasets for polymyxin B, linezolid, and tacrolimus. For polymyxin B and linezolid, repeated subject-wise cross-validation with nested tuning was used to compare the mechanistic baseline with unconstrained and constrained hybrid corrections and a boosted-tree residual benchmark. The studies were designed to assess its behavior in a main application setting, across different drugs, under difficult fitting conditions, and under changes in correction strength and mechanistic parameters. Computational time was also assessed. **Results:** The results showed that the constrained correction remained close to the mechanistic baseline in the held-out analyses of polymyxin B and linezolid, but it did not significantly improve subject-level prediction. The unconstrained correction showed greater deterioration, while the boosted-tree benchmark gave mixed results and no significant subject-level improvement. The additional analyses showed that tighter correction bounds were generally selected and that the constrained hybrid still responds to changes in the mechanistic parameters in a sensible manner. **Conclusions:** Overall, the results suggest that the bounded data-driven correction can control the poorer performance seen with an unrestricted correction while keeping prediction close to the mechanistic baseline. It therefore provides a cautious way to combine mechanistic pharmacokinetic modeling with data-driven correction while preserving interpretability. Further external validation is still needed.

Keywords: compartmental pharmacokinetic models; hybrid pharmacokinetic models; bounded data-driven correction; model validation; residual learning; mechanistic interpretability

Academic Editor: Simone Brogi

Received: 22 June 2026

Revised: 27 July 2026

Accepted: 28 July 2026

Published: 5 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Pharmacokinetic (PK) modeling remains one of the main tools used to describe drug absorption, distribution, and elimination, and it continues to play an important role in dose selection, therapeutic drug monitoring, and individualized treatment planning [1–4]. Classical compartmental and population PK models are still widely used because they are interpretable, physiologically meaningful, and acceptable in clinical and regulatory settings [1,2,5]. However, these models depend on assumptions that may not explain all the variability seen in heterogeneous patient groups or in datasets with limited sampling [2–4]. Therefore, there is still a need for models that keep the mechanistic basis of PK modeling but can also account for patterns that are not explained by the original model. This is important in model-informed precision dosing, where the model should support dose adjustment while remaining understandable in clinical practice [2–4].

Machine-learning and neural differential equation methods have also been used in PK modeling because they can learn patterns that may be missed by standard models [6–9]. However, fully data-driven models may be difficult to interpret, may perform poorly outside the data used for training, and may not preserve basic PK properties such as positivity and compartmental consistency [6–8]. Hybrid models try to address this by combining mechanistic equations with a learned correction [7–9]. Existing hybrid approaches differ in how the learned component is incorporated and controlled. In particular, not all formulations explicitly limit the correction relative to the mechanistic dynamics or examine whether a constrained formulation remains useful under held-out evaluation in different clinical datasets. The present work focuses on this setting by evaluating a bounded correction across published pharmacokinetic datasets with different drugs, dosing structures, and covariate profiles.

This issue is important for drugs with narrow therapeutic ranges or large differences between patients. Tacrolimus, linezolid, and polymyxin B were selected because they represent different clinical settings, including therapeutic drug monitoring, pediatric anti-infective treatment, and treatment of critically ill patients [5,10–14]. The three datasets also differ in their dosing, sampling, and available covariates. They were therefore used to examine whether the framework could be applied across different clinical datasets rather than only in one setting. The central research question of the present study is therefore whether a structure-preserving hybrid PK model can improve or preserve predictive performance under held-out evaluation without losing the mechanistic properties that make PK models useful in practice.

In our earlier work [15], we introduced the general structure-preserving formulation for compartmental PK models, where the learned correction was bounded so that the main mechanistic behavior of the system was not lost. That study established the general formulation and examined its main theoretical and numerical properties. The present study extends this work by examining its performance across published clinical pharmacokinetic datasets, including held-out evaluation in the polymyxin B and linezolid datasets. We therefore applied it to three published clinical datasets involving different drugs, dosing routes, compartmental structures, sampling patterns, and covariate information. We also considered a more difficult case in which the mechanistic baseline was weak. In addition, the present work includes repeated subject-wise cross-validation with nested tuning for polymyxin B and linezolid, comparison with a boosted-tree residual benchmark, further analysis of correction strength, mechanistic parameter sensitivity, computational time, and subject-level performance. Together, these clinical applications and additional analyses form the main contribution of the present study. Unlike many general PK/ADMET or purely data-driven machine-learning tools, the method used in the present study starts from a fitted mechanistic PK model and applies a bounded correction

to its concentration predictions rather than predicting drug properties directly from chemical descriptors.

2. Materials and Methods

2.1. Structure-Preserving Hybrid Pharmacokinetic Framework

We consider a compartmental pharmacokinetic model written in terms of drug amounts. Let $A_i(t)$ denote the amount of drug in compartment i at time t , for $i = 1, \dots, n$, and let $C_i(t) = A_i(t)/V_i$ denote the corresponding concentration, where V_i is the volume of that compartment. The mechanistic model is written as

$$\frac{dA_i}{dt} = F_i(A, \theta, z, u, t), \quad i = 1, \dots, n, \quad (1)$$

where $A = (A_1, \dots, A_n)^T$ is the state vector, θ is the parameter vector, z denotes observed covariates, and u represents known inputs such as dose administration. The functions F_i describe the usual PK processes, including absorption, transfer between compartments, distribution, and elimination. This formulation is sufficiently general for the one- and two-compartment settings considered in the present study.

In the general ODE-based formulation introduced in [15], the hybrid model is obtained by adding a learned correction term to the mechanistic system. Thus, we write

$$\frac{dA_i}{dt} = F_i(A, \theta, z, u, t) + G_i(A, z, u, t; \phi), \quad i = 1, \dots, n, \quad (2)$$

where G_i is a data-driven correction with parameter vector ϕ . The mechanistic component remains central in the formulation. This formulation provides the general basis for the bounded-correction approach used in the present study.

Within this formulation, the correction is controlled by

$$|G_i(A, z, u, t; \phi)| \leq \alpha_i \Psi_i(A, \theta, z, u, t), \quad (3)$$

where $\alpha_i > 0$ is a control parameter and Ψ_i is a suitable reference scale derived from the mechanistic model. This condition limits the correction relative to the main PK dynamics. When an internal transfer term is modified, the correction is constructed to preserve compartmental balance, and it is restricted accordingly when nonnegative state variables are required. These conditions apply to the general ODE-level formulation, while the output-level implementation used in the present numerical studies is described below.

For the numerical studies, the bounded correction was applied to the mechanistic concentration prediction at the output level and expressed on the logarithmic concentration scale. Let $z_\phi(x)$ denote the fitted correction function based on the standardized hybrid features. The constrained correction was defined as $\delta_{\text{con}}(x) = \alpha \tanh(z_\phi(x))$, and the final prediction was calculated as $\hat{C}_{\text{hyb}} = \exp\{\log(\hat{C}_{\text{mech}} + \varepsilon) + \delta_{\text{con}}(x)\}$, where $\varepsilon > 0$ is a small constant used to avoid taking the logarithm of zero. This formulation ensures that $|\delta_{\text{con}}(x)| \leq \alpha$. The correction function included linear terms with or without radial basis functions and was estimated by ridge-regularized least squares. The hyperbolic tangent was used because it provides a smooth and differentiable bound rather than abrupt clipping. For the held-out polymyxin B and linezolid analyses, the value of α was selected during nested inner validation. For illustration, when $\alpha = 0.35$, the correction is limited approximately to a multiplicative range of $\exp(-0.35)$ to $\exp(0.35)$, corresponding to about 0.70 to 1.42 times the mechanistic prediction. Since $|\delta_{\text{con}}(x)| \leq \alpha$, the corrected prediction satisfies $e^{-\alpha}(\hat{C}_{\text{mech}} + \varepsilon) \leq \hat{C}_{\text{hyb}} \leq e^{\alpha}(\hat{C}_{\text{mech}} + \varepsilon)$. Therefore, a positive mechanistic prediction remains positive after correction, and the learned component cannot move the prediction beyond the prescribed multiplicative range. The smoothness of the hyperbolic tangent also avoids discontinuities in the corrected prediction.

2.2. Population Pharmacokinetic Setting and Estimation

The present work is carried out in a population-oriented PK setting, in which parameters may vary across individuals through observed covariates and, in the general formulation, random effects. At the structural level, this is written as

$$\theta = \theta(z, \eta), \quad (4)$$

where z contains observed covariates and η denotes interindividual variability. In this way, the mechanistic model retains the main PK interpretation, while the learned correction is used only for residual systematic effects that remain after the structural model and covariate relations have been introduced. Parameters such as clearance, volume, and absorption rate therefore continue to play their usual role in the formulation. In the present numerical studies, subject-specific parameters were represented using the available covariate relations, and a full nonlinear mixed-effects estimation of random effects was not performed.

Let Y_{ij} denote the measured concentration for subject j at time t_{ij} . The observation model is written as

$$Y_{ij} = \mathcal{H}(A_j(t_{ij})) + \varepsilon_{ij}, \quad (5)$$

where $\mathcal{H}$ maps the compartmental state to the observed quantity and ε_{ij} denotes residual error. Parameter estimation is based on an objective function that combines data fit with control of the correction term. For the output-level implementation used in the numerical studies, let δ denote the correction applied to the mechanistic concentration prediction. A general form of the objective function is

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda \mathcal{R}(\delta), \quad (6)$$

where $\mathcal{L}_{\text{data}}$ measures the discrepancy between observed and predicted concentrations, $\mathcal{R}(\delta)$ penalizes excessive correction size or behavior inconsistent with the intended structure, and $\lambda > 0$ is a regularization parameter. The exact residual model and regularization term may differ across drugs, but the same general framework is used throughout the study.

In the numerical studies, the mechanistic parameters were estimated using nonlinear least squares. For the held-out polymyxin B and linezolid analyses, the mechanistic parameters were re-estimated using only the training subjects within each fold. The correction model combined standardized linear features with or without radial basis functions and was fitted using ridge-regularized least squares. The correction form, regularization level, and correction-strength parameter were selected using nested inner validation, as described below. The correction parameters were obtained by solving the regularized linear system directly. Convergence of the mechanistic fitting was assessed using the predefined function and step tolerances.

2.3. Mathematical Properties of the Hybrid Model

The analysis below concerns the general ODE-based hybrid formulation in (2), together with the control condition in (3), the population parameter relation in (4), and the observation structure in (5) and (6). We assume that the functions appearing in (1) and (2) are continuous in time, continuously differentiable with respect to the state and parameter variables, and continuous in the remaining variables on the region of interest.

Proposition 1. *Assume that the vector field defined by the right-hand side of (2) is locally Lipschitz in the state variable, uniformly with respect to the parameters and patient descriptors on bounded sets. Then, for every initial condition, the hybrid system admits a unique local solution. Moreover, the solution depends continuously on the initial data, the mechanistic parameters, the correction parameters, and the patient descriptors appearing in (4).*

Small changes in the patient covariates or parameter values do not lead to abrupt changes in the PK trajectory, at least on finite time intervals where the solution exists. This is important in the present setting because the same framework is used across different patients and different clinical datasets. A proof is given in Appendix A. The result is a standard consequence of local Lipschitz continuity, but here it is used in a population-oriented PK setting where the model depends explicitly on patient descriptors [16].

Proposition 2. *Let $S_\theta(t)$ denote the sensitivity matrix of the hybrid model with respect to the mechanistic parameters, and let $S_\theta^0(t)$ denote the corresponding sensitivity matrix for the mechanistic model in (1). Assume that the derivatives of the correction term with respect to the state and, if present, with respect to the mechanistic parameters are bounded on the region of interest, and that their combined size is at most $\varepsilon > 0$. Then, on every finite time interval on which both solutions remain in that region, there exists a constant $C_T > 0$, depending on the interval and on bounds of the mechanistic system but independent of ε , such that*

$$\|S_\theta(t) - S_\theta^0(t)\| \leq C_T \varepsilon$$

If the correction remains sufficiently controlled, then the way the hybrid model responds to changes in the mechanistic parameters stays close to the response of the base PK model. This matters because quantities such as clearance, volume, and absorption rate should still retain their main PK meaning after the correction is introduced. This behavior was discussed qualitatively in [15], while Proposition 2 gives an explicit finite-time bound under the stated derivative condition. This result follows from the variational equations associated with (1) and (2), together with the stated derivative bound on the correction term and a Grönwall-type estimate. The proof is included in Appendix A.

Proposition 3. *Suppose that, for the mechanistic model observed through (5), the parameter-to-observation map is locally distinguishable at a reference parameter value, in the sense that the corresponding sensitivity matrix has full column rank on the observation window of interest. Assume further that the observation map H is continuously differentiable, and that the correction term is continuously differentiable and satisfies the smallness condition used in Proposition 2. Then there exists $\varepsilon_0 > 0$ such that, whenever that smallness condition holds with threshold ε_0 , the hybrid model preserves local distinguishability of the mechanistic parameters on the same observation window.*

If the mechanistic parameters can be locally separated in the base model, then a sufficiently controlled correction does not destroy that property. In practical terms, the main PK parameters remain locally distinguishable as long as the learned term does not compete too strongly with them. This point was discussed qualitatively in [15], while Proposition 3 gives a local rank-based result under the stated smallness condition. This proposition does not claim full structural identifiability of the joint parameter set. It is a local result for the mechanistic parameters in the presence of a constrained correction. The argument relies on continuity of matrix rank under sufficiently small perturbations and is given in Appendix A.

In the general model, if a correction is added to a transfer term, it is added with opposite signs in the two compartments. In this way, the correction only changes the movement of the drug between compartments and does not add or remove drug from the system. Equation (3) also keeps the size of this correction limited relative to the mechanistic model. These statements refer to the general ODE-level formulation.

For the output-level implementation used in the numerical analyses, positivity and bounded correction size follow directly from the form of the corrected prediction.

Propositions 1–3 concern the general ODE-based formulation and provide the analytical basis for its population-oriented extension.

2.4. Datasets and Numerical Implementation

We tested the framework using three published clinical pharmacokinetic datasets. The datasets were selected because they contained concentration-time data together with dosing information and relevant patient covariates, and because they represented different drugs, dosing routes, and model structures. Each dataset was analyzed separately and the datasets were not combined into a single training dataset. Repeated subject-wise validation was used for polymyxin B and linezolid, whereas tacrolimus was evaluated as an exploratory in-sample robustness study. No independent external validation dataset was used. The numerical studies are organized by purpose. We begin with a primary clinical application on polymyxin B, followed by cross-drug evaluation and a robustness study in a more difficult clinical setting. Only records with a valid observed concentration were included in the analysis. Missing concentration values were excluded and were not imputed. We first consider the polymyxin B dataset from Hanafin et al. [13]. This dataset was chosen as the first main application because it gives a clinically important setting with explicit intravenous infusion records and a mechanistic structure that is suitable for a full comparison. In our implementation, the dataset contained 40 subjects, 268 dosing events, and 230 concentration observations. The data was provided in event format, with explicit entries for dose amount, infusion rate, and infusion duration. Therefore, in this case we did not need to reconstruct oral absorption, lag time, or repeated dosing from simplified summaries. The infusion events were taken directly from the file and used as they were. For the mechanistic baseline, we used a two-compartment intravenous infusion model with elimination from the central compartment. The numerical model was

$$\begin{aligned}\frac{dA_1}{dt} &= R_{\text{in}}(t) - \left(\frac{CL_i}{V_{c,i}} + \frac{Q_i}{V_{c,i}} \right) A_1 + \frac{Q_i}{V_{p,i}} A_2, \\ \frac{dA_2}{dt} &= \frac{Q_i}{V_{c,i}} A_1 - \frac{Q_i}{V_{p,i}} A_2,\end{aligned}$$

where A_1 and A_2 are the drug amounts in the central and peripheral compartments, respectively, and $R_{\text{in}}(t)$ is the infusion input rate. The subject-specific parameters were scaled by body weight, and clearance also included a creatinine clearance effect. In the final mechanistic baseline, the predicted concentration was taken as

$$\hat{C}_{\text{mech}}(t) = \frac{A_1(t)}{V_{c,i}} + C_0.$$

Among the mechanistic runs we tested, this offset version gave the best baseline fit and was therefore retained as the mechanistic model structure in this application.

We next considered linezolid as a cross-drug evaluation study using the published dataset from [12], where the purpose was to assess the held-out predictive performance of the framework in a different pharmacokinetic setting. In this case, the data corresponded to oral or nasogastric administration, so the mechanistic baseline was simpler than in the polymyxin B study. After data cleaning and dose expansion, the dataset used in our implementation contained 48 subjects, 305 dose events, and 196 concentration observations.

For the mechanistic baseline, we used a one-compartment model with first-order absorption and first-order elimination. The numerical model was

$$\frac{dA_g}{dt} = -K_a A_g, \quad \frac{dA_c}{dt} = K_a A_g - \frac{CL_i}{V_i} A_c,$$

where A_g and A_c denote the amounts in the absorption and central compartments, respectively. The predicted concentration was taken as

$$\hat{C}_{mech}(t) = \frac{A_c(t)}{V_i}.$$

The subject-specific parameters were scaled by body weight, and the mechanistic parameters were re-estimated using only the training subjects within each outer fold.

We next considered the published tacrolimus dataset of Quintairos et al. as an exploratory robustness study [17]. The role of this experiment was different from the first two studies. Here, the goal was to test whether the framework could reduce in-sample prediction error when the baseline model is weak and the data are more difficult. This is a harder setting because tacrolimus is well known for large variability, and in our implementation the data were naturally handled at the occasion level using the observed time-after-dose values. After preprocessing, the dataset used in this study contained 471 occasions and 1102 concentration observations. For the mechanistic baseline, we used an oral two-compartment model with first-order absorption, lag time, and elimination from the central compartment. The numerical model was

$$\begin{aligned}\frac{dA_g}{dt} &= -K_a A_g, \\ \frac{dA_c}{dt} &= K_a A_g - \left(\frac{CL_i}{V_{c,i}} + \frac{Q_i}{V_{c,i}} \right) A_c + \frac{Q_i}{V_{p,i}} A_p, \\ \frac{dA_p}{dt} &= \frac{Q_i}{V_{c,i}} A_c - \frac{Q_i}{V_{p,i}} A_p,\end{aligned}$$

where A_g , A_c , and A_p denote the amounts in the absorption, central, and peripheral compartments, respectively. The predicted concentration was taken as

$$\hat{C}_{mech}(t) = \frac{A_c(t)}{V_{c,i}}.$$

Across all three numerical studies, the correction was applied to the mechanistic concentration prediction at each observation time. The hybrid component did not directly estimate or modify pharmacokinetic parameters such as clearance, volume of distribution, intercompartmental clearance, absorption rate, or lag time. These parameters remained part of the fitted mechanistic baseline. The main predicted endpoint was drug concentration, and model performance was assessed by comparing the predicted and observed concentrations. For polymyxin B and linezolid, predictive performance was assessed in held-out subjects. For tacrolimus, the reported comparison was exploratory and in-sample.

2.5. Model Comparison and Sensitivity Analyses

Whenever relevant, we compare the mechanistic baseline with unconstrained and constrained hybrid models and, for polymyxin B and linezolid, with a boosted-tree residual benchmark to examine whether the correction improves or preserves held-out predictive performance while remaining controlled relative to the mechanistic baseline. Model performance was assessed using the root mean square error (RMSE), mean absolute error (MAE), median absolute error (MedAE), correlation coefficient, and logarithmic root mean square error (LogRMSE).

For the polymyxin B and linezolid analyses, we used repeated five-fold subject-wise cross-validation with five repeats. All records from the same subject were kept within the same fold. In each repeat, the mechanistic model and the correction models were fitted using the outer training subjects and evaluated using only the held-out subjects. Within each outer training set, four-fold subject-wise inner validation was used for model selection. The mechanistic parameters were re-estimated using only the inner training subjects.

Any covariate imputation, feature standardization, and radial basis function center selection were also based only on the corresponding training data.

The candidate correction models included a linear-only form and radial basis function forms with 4, 8, or 12 centers. Ridge regularization values of 0.01, 0.1, 1, and 10 were tested. For the constrained correction, values $\alpha \in \{0.10, 0.15, 0.25, 0.35\}$ were considered. This gave 16 unconstrained and 64 constrained candidates. The unconstrained and constrained models were selected separately according to the mean subject-level MAE in the inner validation folds. After selection, the correction was fitted using the complete outer training set and evaluated using the held-out outer subjects.

The boosted-tree residual benchmark was fitted to the same log-scale residual target and standardized features used by the correction models. It used least-squares gradient boosting with 150 regression stumps, a learning rate of 0.05, and a minimum leaf size of 5 observations. These settings were fixed before the outer held-out evaluation.

The effect of the correction-strength parameter was examined through its selection during nested inner validation. The frequency with which each value of α was selected across the 25 outer folds was recorded separately for polymyxin B and linezolid. No single value of α was fixed for the held-out analyses.

We finally checked whether the constrained hybrid model still responds to changes in the main mechanistic parameters in a sensible way. This test was done on the polymyxin B dataset, since this was the main application and gave the clearest baseline model. The aim here was to see whether the hybrid correction keeps the main effect of the mechanistic parameters or hides it completely. We considered three parameters from the mechanistic model, namely CL_{70} , $V_{C,70}$, and Q_{70} , and perturbed each of them by $\pm 10\%$ while keeping the remaining parts of the model fixed. This separate analysis used $\alpha = 0.35$ and was treated as a local in-sample interpretability analysis rather than a held-out predictive comparison.

Computational time was assessed for the polymyxin B application using one unrecorded warm-up run followed by three measured runs on the same computer. The mechanistic prediction, feature construction, correction fitting, bounded transformation, and final hybrid prediction were timed separately. Data loading, file writing, and figure generation were excluded. The timing results are reported as mean $\pm$ standard deviation. This timing comparison concerned one mechanistic fit and one correction fit and did not include the repeated cross-validation or nested model-selection procedure.

For each outer repeat, the held-out predictions from the five outer folds were combined before the overall performance metrics were calculated. These metrics were then summarized across the five repeats using the mean and standard deviation. To examine performance across subjects, the MAE was calculated separately for each subject using its held-out predictions and then averaged across the five repeats. Paired subject-level differences were defined as the mechanistic MAE minus the comparator MAE, so that a positive value indicated improvement over the mechanistic model. The paired values were compared using a two-sided Wilcoxon signed-rank test, and the rank-biserial effect size was calculated. Ninety-five percent confidence intervals for the mean and median paired differences were estimated using 10,000 subject-level bootstrap samples. The constrained-versus-mechanistic comparison was the main comparison, while the unconstrained and boosted-tree comparisons were additional comparisons. A p -value below 0.05 was considered statistically significant.

3. Results

3.1. Primary Clinical Application: Polymyxin B

For the repeated held-out polymyxin B analysis, the mechanistic model was re-estimated within each training fold and compared with two tuned hybrid corrections and a boosted-tree residual benchmark. The first was an unconstrained hybrid model, where a learned correction was added on the logarithmic scale to the mechanistic prediction. The second was a constrained hybrid model, where a correction of the same form was bounded smoothly so that it remained controlled relative to the mechanistic baseline. Thus, four models were compared in this study, the mechanistic baseline, the tuned unconstrained hybrid, the tuned constrained hybrid, and the boosted-tree residual benchmark.

The main results are summarized in Table 1. The mechanistic baseline gave the best overall held-out predictive metrics. The mean MAE across the five repeats was 1.8788 for the mechanistic model, 2.1805 for the tuned unconstrained hybrid, 1.9427 for the tuned constrained hybrid, and 2.0742 for the boosted-tree benchmark. Among the data-driven models, the constrained hybrid remained closest to the mechanistic baseline and showed less deterioration than the unconstrained correction. This is the main idea of the bounded-correction approach used in the present study. The goal is not only to reduce the prediction error, but also to keep the learned correction under control and preserve the role of the mechanistic model. The held-out results therefore support the value of the constraint in controlling residual learning, although they do not show an overall predictive improvement over the mechanistic model. At the subject level, the constrained hybrid had a lower mean absolute error in 16 of the 40 subjects. The median subject-level mean absolute error was 1.7895 for the mechanistic model and 1.9146 for the constrained hybrid. The difference did not reach statistical significance in the two-sided Wilcoxon signed-rank test ($p = 0.106$). The complete patient-level paired results, including the bootstrap confidence intervals and rank-biserial effect sizes, are reported in Table A1. In comparison, the unconstrained hybrid was significantly worse than the mechanistic baseline at the subject level ($p = 0.007$).

Representative held-out subject profiles are shown in Figure 1. The mechanistic model captures the general trend, while the constrained hybrid generally remains close to the mechanistic prediction. The profiles show that the correction can reduce part of the mismatch in some subjects but does not do so consistently across all subjects. Together with the results in Table 1, this indicates that the constrained correction limits the larger changes seen with unrestricted residual learning without demonstrating overall superiority to the mechanistic baseline.

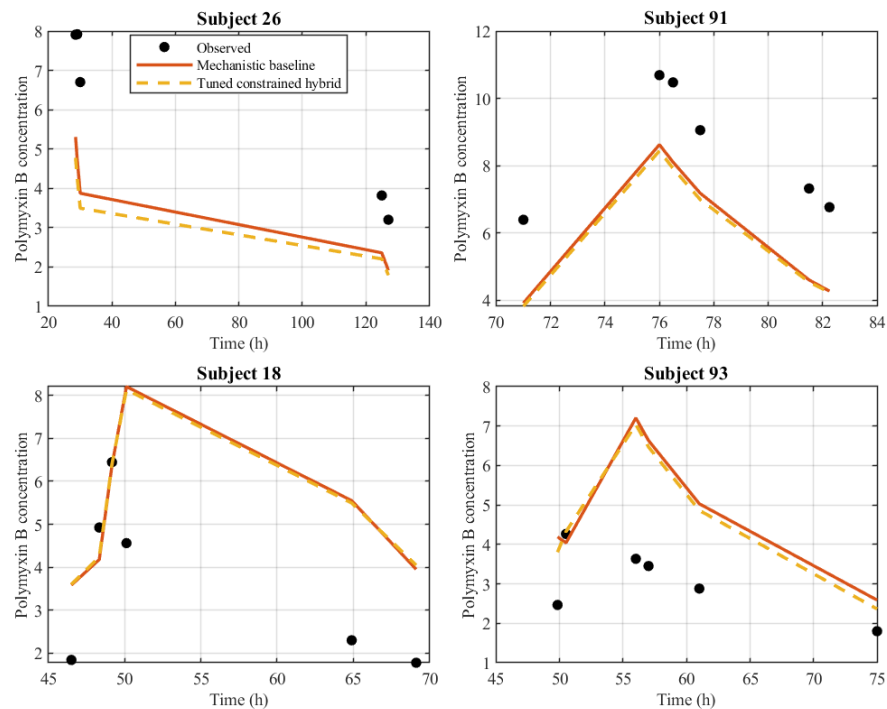


Figure 1. Representative held-out concentration–time profiles for four subjects in the polymyxin B application. Subjects were selected from different quartiles of the subject-level difference in mean absolute error between the mechanistic and constrained models.

Table 1. Held-out predictive performance in the primary polymyxin B application using repeated subject-wise cross-validation and nested model selection. Values are the mean (SD) across five repeats.

Model	RMSE	MAE	MedAE	Correlation	LogRMSE
Mechanistic	2.432 (0.040)	1.879 (0.036)	1.536 (0.028)	0.712 (0.011)	0.482 (0.013)
Unconstrained hybrid	3.032 (0.423)	2.181 (0.225)	1.661 (0.170)	0.600 (0.066)	0.563 (0.072)
Constrained hybrid	2.516 (0.060)	1.943 (0.070)	1.599 (0.107)	0.695 (0.017)	0.499 (0.015)
Boosted-tree benchmark	2.707 (0.214)	2.074 (0.139)	1.684 (0.108)	0.679 (0.019)	0.517 (0.027)

3.2. Cross-Drug Evaluation: Linezolid

For the repeated held-out linezolid analysis, the mechanistic model was re-estimated within each training fold and compared with the tuned unconstrained and constrained hybrid models and the boosted-tree residual benchmark. The main results are summarized in Table 2. The tuned constrained hybrid remained close to the mechanistic baseline, while the unconstrained correction showed poorer overall held-out performance.

Table 2. Held-out predictive performance in the linezolid evaluation using repeated subject-wise cross-validation and nested model selection. Values are mean (SD) across five repeats.

Model	RMSE	MAE	MedAE	Correlation	LogRMSE
Mechanistic	3.876 (0.027)	2.767 (0.019)	1.785 (0.052)	0.647 (0.008)	0.569 (0.003)
Unconstrained hybrid	4.272 (0.581)	2.937 (0.234)	1.839 (0.040)	0.588 (0.039)	0.587 (0.018)
Constrained hybrid	3.937 (0.055)	2.782 (0.054)	1.908 (0.073)	0.615 (0.011)	0.582 (0.008)
Boosted-tree benchmark	3.755 (0.093)	2.683 (0.051)	1.847 (0.071)	0.673 (0.019)	0.550 (0.008)

The mean held-out MAE across the five repeats was 2.7667 for the mechanistic model, 2.9371 for the tuned unconstrained hybrid, 2.7824 for the tuned constrained hybrid, and 2.6832 for the boosted-tree benchmark. The constrained hybrid therefore remained close

to the mechanistic baseline and performed better overall than the unconstrained correction, but it did not provide a clear improvement over the mechanistic model.

At the subject level, the constrained hybrid had a lower MAE in 26 of the 48 subjects. The median subject-level MAE was 2.1177 for the mechanistic model and 2.2410 for the constrained hybrid. The paired difference was not statistically significant in the two-sided Wilcoxon signed-rank test ($p = 0.867$). The boosted-tree benchmark gave the lowest overall mean MAE, but its subject-level difference from the mechanistic model was also not statistically significant ($p = 0.467$). The complete patient-level paired results are reported in Table A2.

Representative held-out subject profiles are shown in Figure 2. The constrained hybrid generally remains close to the mechanistic prediction, with small changes in either direction across subjects. Together with the results in Table 2, this indicates that the bounded correction limited the poorer performance of the unconstrained correction but did not consistently improve prediction over the mechanistic baseline.

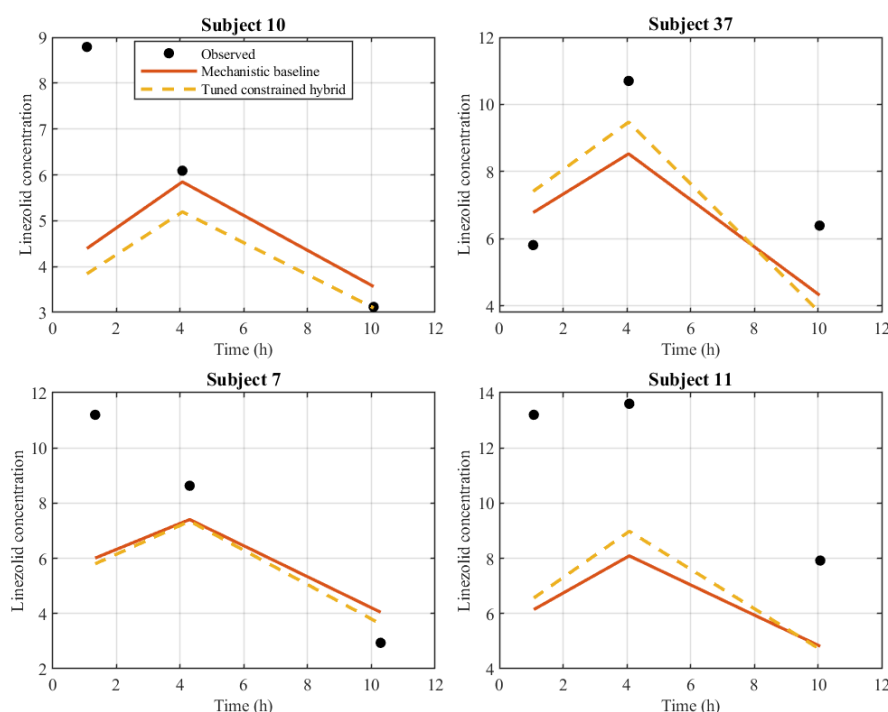


Figure 2. Representative held-out concentration–time profiles for four subjects in the linezolid evaluation. Predictions were averaged across the five repeated out-of-fold evaluations.

3.3. Exploratory Application in a Difficult Clinical Setting: Tacrolimus

The mechanistic run showed that this was a difficult baseline. The fitted parameters moved close to the imposed bounds, and the resulting predictions were systematically low relative to the observations. This is seen clearly in Figure 3, where the mechanistic curves capture only the broad shape and underpredict the measured concentrations over most of the dosing interval.

After fixing this weak baseline, we applied the constrained hybrid correction. The purpose here was to see whether a bounded correction could reduce the in-sample prediction error in a difficult setting while keeping the mechanistic baseline as the basis of the prediction. The results are summarized in Table 3. The constrained hybrid reduced all main error measures relative to the mechanistic baseline. The gain in correlation was small, but the reduction in absolute error and log-scale error was clear. This is the main point of the tacrolimus study. Within this exploratory in-sample analysis, the constrained

correction moved the predictions in the right direction and reduced the strong underprediction bias. In this sense, the experiment serves as an exploratory difficult-fitting application rather than as a claim of a high-quality tacrolimus PK model or evidence of held-out predictive performance.

Table 3. In-sample summary metrics for the exploratory tacrolimus application.

Model	RMSE	MAE	MedAE	Correlation	LogRMSE
Mechanistic baseline	2.3057	2.2341	2.2566	0.4631	1.9712
Hybrid constrained	2.0422	1.9557	1.9882	0.4632	1.5491

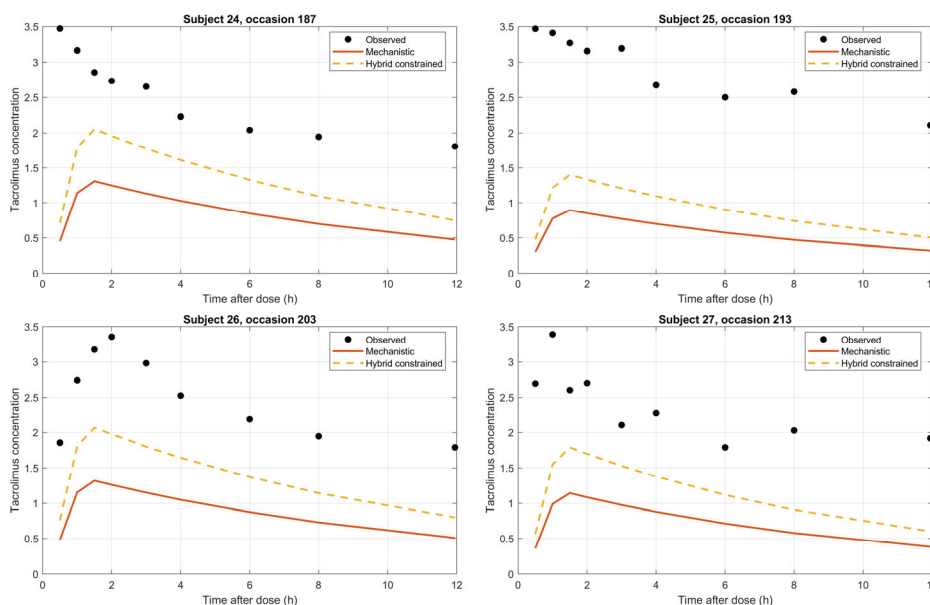


Figure 3. Representative occasion-level concentration–time profiles for four tacrolimus cases in the exploratory in-sample application.

Representative occasion-level profiles are shown in Figure 3. The constrained hybrid pushes the predictions upward and follows the observed levels better than the mechanistic baseline, while still preserving the same general profile shape. Together with the summary in Table 3, this shows that the correction reduced in-sample error in a substantially more difficult clinical setting.

3.4. Selection of Correction Strength

The correction-strength selection results are summarized in Table 4. For polymyxin B, $\alpha=0.10$ was selected in 20 of the 25 outer folds, while $\alpha = 0.15$ was selected in the remaining five folds. For linezolid, $\alpha = 0.10, 0.15$, and 0.25 were selected in 7, 12, and 6 folds, respectively. The value $\alpha = 0.35$ was not selected in either dataset. These results show that nested inner validation generally favored tighter correction bounds, particularly in the polymyxin B application. The selected values also differed between the two datasets, supporting validation-based selection of the correction strength rather than the use of one fixed value across different settings.

The bounded correction introduced only a small computational overhead. The fixed mechanistic prediction required 0.40285 ± 0.00300 s, while the complete constrained-hybrid step added 0.00555 ± 0.00305 s, corresponding to approximately 1.38% additional computational time. The target transformation and application of the bound together required only 0.000532 ± 0.000378 s. For comparison, fitting and predicting the full

mechanistic model required 75.078 ± 0.168 s. These values concern a single mechanistic and correction fit and do not include repeated cross-validation or nested model selection.

Table 4. Frequency of correction-strength values selected during nested inner validation across the 25 outer folds for polymyxin B and linezolid.

α	Polymyxin B, n (%)	Linezolid, n (%)
0.10	20 (80%)	7 (28%)
0.15	5 (20%)	12 (48%)
0.25	0 (0%)	6 (24%)
0.35	0 (0%)	0 (0%)

3.5. Mechanistic Parameter Sensitivity and Interpretability

The results of the local in-sample mechanistic parameter sensitivity study are summarized in Table 5, and the main trends are shown in Figures 4 and 5. The same pattern appears in all cases. The constrained hybrid still changes when the mechanistic parameters are perturbed, but the size of this change is smaller than in the pure mechanistic model. This means that the hybrid correction does not remove the role of the mechanistic parameters, but it reduces the sensitivity of the final prediction to small changes in them.

At the same time, the relative importance of the parameters remains reasonable. In this experiment, perturbations in CL_{70} produced the largest change, perturbations in $V_{C,70}$ gave a moderate change, and perturbations in Q_{70} gave the smallest one. So the main mechanistic ordering was kept. The mean relative prediction changes reported in Table 5 support the same conclusion. The mechanistic model changed more across the perturbation cases, whereas the constrained hybrid stayed in a narrower range.

Representative concentration–time profiles for a +10% perturbation in CL_{70} are shown in Figure 5. Both the mechanistic and hybrid predictions move in the expected direction, but the hybrid does not react sharply and still keeps the overall profile shape. So the main conclusion of this experiment is that the constrained hybrid keeps the main mechanistic influence structure, while reducing the magnitude of the sensitivity in the final predictions.

Table 5. Local in-sample sensitivity of the mechanistic and constrained hybrid predictions to $\pm 10\%$ perturbations in selected mechanistic parameters for the polymyxin B dataset. The constrained hybrid used $\alpha = 0.35$.

Perturbation Case	Mechanistic Mean Relative Prediction Change	Hybrid Mean Relative Prediction Change
$CL_{70} + 10\%$	0.0598	0.0449
$CL_{70} - 10\%$	0.0693	0.0502
$Q_{70} + 10\%$	0.0128	0.0092
$Q_{70} - 10\%$	0.0144	0.0103
$V_{C,70} + 10\%$	0.0293	0.0214
$V_{C,70} - 10\%$	0.0344	0.0254

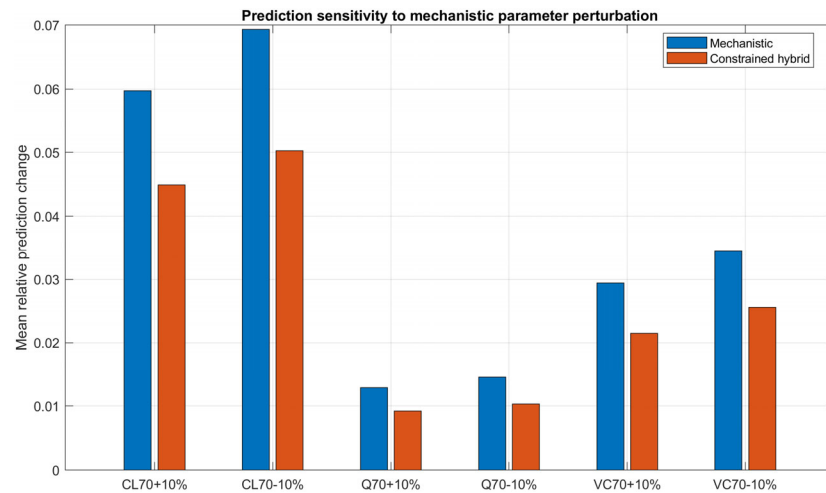


Figure 4. Mean relative prediction change under $\pm 10\%$ perturbations in selected mechanistic parameters in the local in-sample analysis.

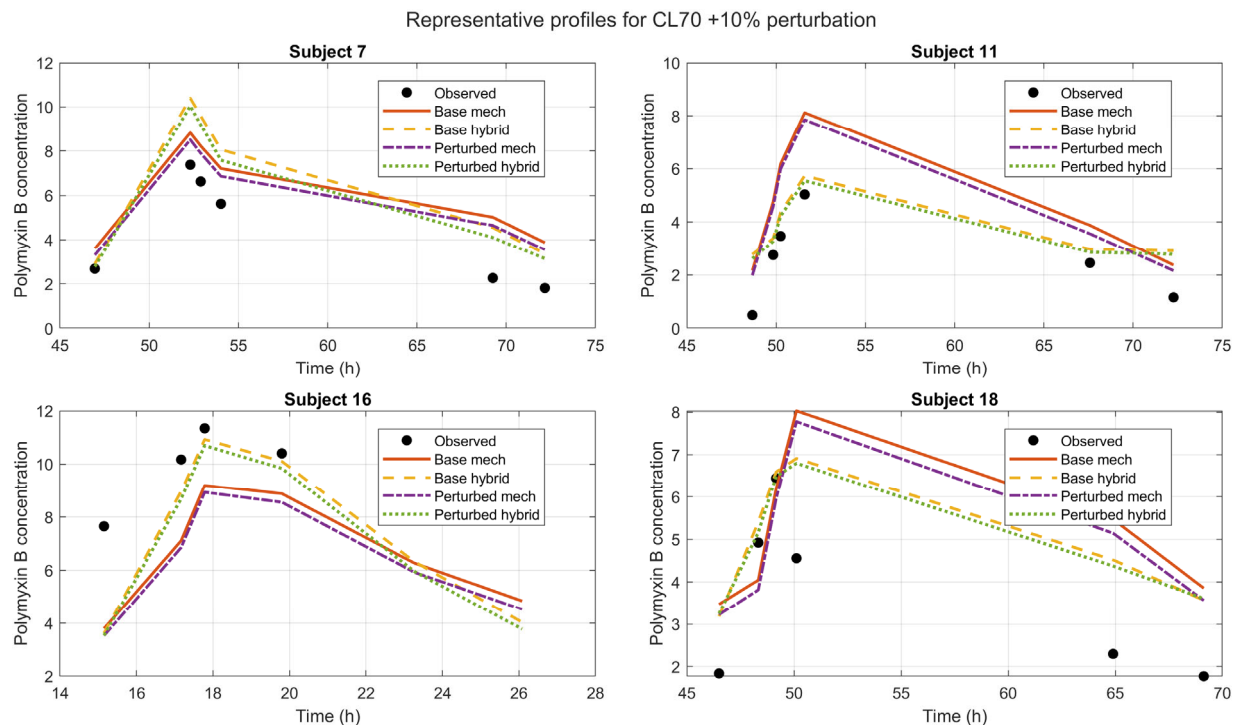


Figure 5. Representative concentration–time profiles under a +10% perturbation in CL_{70} in the local in-sample analysis.

4. Discussion

The numerical results support the main idea of the framework. The present study extends the earlier ODE-based work in [15] by evaluating the bounded correction at the output level across three clinical datasets. In the polymyxin B study, the constrained hybrid remained close to the mechanistic baseline under held-out evaluation, while the unconstrained correction showed greater deterioration. In the linezolid study, the constrained hybrid again remained close to the mechanistic model. The boosted-tree benchmark gave the lowest overall mean MAE in this dataset, although its subject-level difference from the mechanistic model was not statistically significant. In the tacrolimus study, the baseline mechanistic model performed poorly, but the constrained hybrid was still

able to reduce the main in-sample prediction errors. We do not claim that this represents a strong PK model for tacrolimus. It was used as a harder exploratory case to check whether the framework still helps when the mechanistic part is weak. Overall, these results suggest that the bounded correction can provide better control than an unrestricted data-driven correction while keeping the mechanistic model central [2–4,7–9]. However, the constrained hybrid did not show significant subject-level improvement over the mechanistic model in either the polymyxin B or linezolid analysis. These findings should therefore be confirmed using larger independent datasets.

The constrained formulation also behaved as intended. Under held-out evaluation, the constrained correction remained closer to the mechanistic baseline than the unconstrained correction, particularly in the polymyxin B study. This was seen clearly in the polymyxin B results and was also supported by the nested selection of the correction-strength parameter. Inner validation generally favored tighter correction bounds. The value $\alpha = 0.35$ was not selected in either the polymyxin B or linezolid analysis, while values between 0.10 and 0.25 were selected across the outer folds. This is useful because it shows that the constraint is not only a formal restriction but also affects the numerical performance of the model. Thus, even when the fitted correction becomes large, the final concentration remains positive and is restricted to the multiplicative range determined by the selected value of α . The last experiment also showed that the constrained hybrid still responds to perturbations in the mechanistic parameters in a sensible way. The main ordering of parameter influence was preserved, while the final predictions became less sensitive than in the pure mechanistic model. Although this was a local in-sample analysis using $\alpha = 0.35$, it supports the view that the correction does not remove the main mechanistic role of the parameters but modifies their effect in a controlled way [2,4,7,15].

The same idea could also be used with nonlinear or larger multi-compartment models because the correction is not tied to one specific PK structure. It may also be extended to full population PK or physiologically based pharmacokinetic models. However, these extensions would require more care in deciding where the correction is added and in preserving positivity, mass balance, and physiological meaning. Random effects, covariate relationships, parameter identifiability, and the larger computational cost would also need to be considered.

The present work also has limitations. The hybrid correction was tested only on the available retrospective datasets, so the results still depend on the size, quality, and variability of those data. Repeated subject-wise validation was carried out for polymyxin B and linezolid, but no independent external dataset was available. The tacrolimus application remained exploratory and in-sample. The tacrolimus study makes this point clear. When the baseline model is weak and the data are challenging, the hybrid may improve the fit, but it cannot fix every limitation in the model. The effect of the initial mechanistic parameter values was not examined in a separate sensitivity study. Although the correction parameters were obtained directly by ridge-regularized least squares, the mechanistic baseline may still depend on the selected starting values and parameter bounds. The learned correction may also be less stable when observations are limited or highly variable. We did not perform a separate analysis for sparse sampling or added measurement noise. Therefore, the performance of the framework under very sparse or noisy data still needs to be examined. The bounded correction can limit large adjustments, but it cannot replace information that is not present in the data. The method may also give little or no improvement when the mechanistic model already fits the data well, when the selected features do not capture the remaining error, or when the correction bound is chosen too small or too large. Repeated cross-validation reduces the risk of an overly optimistic result, but it does not replace external or prospective validation.

As the aim of this work was to test whether a structure-preserving hybrid correction can improve or preserve the predictive performance of a mechanistic PK backbone while preserving its main interpretation, we focused on that objective directly. A full population PK re-estimation with random effects, complete covariate model development, and external validation in the conventional pharmacometric sense were therefore beyond the scope of the present paper. Within that scope, the results are mixed but informative. They show that the bounded correction does not guarantee improved prediction, but that tighter control can limit the deterioration seen with unrestricted residual learning. Larger external studies and prospective validation are still needed before such models can be used in routine clinical dose individualization [2,4,13,14]. The three drugs also belong to different chemical and pharmacological classes. Tacrolimus is a macrolide immunosuppressant, linezolid is an oxazolidinone antibiotic, and polymyxin B is a cyclic lipopeptide antibiotic. However, the study includes only three compounds and was not designed as a formal chemical-space analysis. The findings therefore demonstrate evaluation across different drug classes and pharmacokinetic settings, but they should not be interpreted as covering the wider chemical space of therapeutic compounds.

A direct numerical comparison with a boosted-tree residual benchmark was carried out using the same held-out folds, residual target, and preprocessing conditions. The benchmark had the lowest aggregate mean MAE in the linezolid analysis, but it did not show a significant subject-level advantage over the mechanistic model. In the polymyxin B analysis, it performed worse than the mechanistic baseline. This comparison therefore did not show consistent superiority of the machine-learning benchmark across the two datasets. A broader comparison with other published hybrid, machine-learning, or physics-informed PK methods was not carried out because these approaches were developed using different datasets, model structures, and fitting procedures. A fair comparison would require their reimplementations using the same data and validation conditions.

Recent deep-learning and multimodal methods show how high-dimensional information from medical images, molecular structures, protein sequences, and biological interaction data can be used in biomedical prediction [18,19]. These developments provide a possible direction for extending the present framework. When suitable data are available, the bounded correction could include molecular, omics, imaging, or other biological features. These extensions could then be compared with the current model using the same datasets and validation conditions.

5. Conclusions

In this work, we evaluated a population-oriented application of the structure-preserving hybrid framework for compartmental pharmacokinetic models in several numerical studies. The results showed that the method did not consistently improve held-out prediction, but the constrained formulation remained closer to the mechanistic baseline than the unconstrained correction while keeping the mechanistic model as the basis of the prediction. In the polymyxin B study, the constrained hybrid remained close to the baseline mechanistic model, while the unconstrained correction showed greater deterioration. In the linezolid study, the same approach again preserved performance close to the mechanistic model, although the boosted-tree benchmark gave the lowest aggregate mean MAE. In the tacrolimus study, the mechanistic baseline was weak, but the constrained correction still helped reduce the main in-sample prediction errors. The additional analyses showed that nested validation generally selected tighter correction bounds, and that the constrained hybrid still responds to changes in the mechanistic parameters in a sensible way. Overall, the results suggest that this framework can combine mechanistic PK structure with data-driven correction without turning the model into a fully black-box predictor. The main value of the bounded formulation was therefore its control of residual

learning rather than consistent predictive superiority. This makes it a potentially useful approach for further hybrid pharmacokinetic modeling research, although external and prospective validation is still needed before clinical use.

Author Contributions: Conceptualization, H.A.L. and M.A.-L.; methodology, M.A.-L. and H.A.L.; software, M.A.-L.; validation, H.A.L., A.A.L. and M.A.-L.; formal analysis, M.A.-L.; investigation, H.A.L., A.A.L. and M.A.-L.; resources, H.A.L. and A.A.L.; data curation, H.A.L. and M.A.-L.; writing—original draft preparation, H.A.L. and M.A.-L.; writing—review and editing, H.A.L., A.A.L. and M.A.-L.; visualization, M.A.-L.; supervision, M.A.-L.; project administration, H.A.L. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank the Oman College of Health Sciences (OCHS) for covering the article processing charges (APC) for this manuscript.

Data Availability Statement: No new experimental or clinical data were generated in this study. The numerical results were obtained using the mathematical models described in this manuscript together with previously published pharmacokinetic datasets cited in the article. The processed data and computational outputs used for the reported figures and tables are available from the corresponding author upon reasonable request due to institutional data-management policies.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PK	Pharmacokinetics
MIPD	Model-informed precision dosing
TDM	Therapeutic drug monitoring
ML	Machine learning
AI	Artificial intelligence
ODE	Ordinary differential equation
RMSE	Root mean square error
MAE	Mean absolute error
MedAE	Median absolute error
LogRMSE	Log-scale root mean square error
CL	Clearance
V	Volume of distribution
V _C	Central compartment volume
V _P	Peripheral compartment volume
Q	Intercompartmental clearance
K _a	Absorption rate constant
C ₀	Concentration offset

Appendix A

This appendix gives the proofs of the analytical results stated in the main text. We use the notation introduced in Sections 2 and 3. In particular, the mechanistic model is given by (1), the hybrid model by (2), and the correction bound by (3).

Proof of Proposition A1. Let

$$H(A, \theta, \phi, z, u, t) = F(A, \theta, z, u, t) + G(A, z, u, t; \phi)$$

denote the right-hand side of (2). By assumption, H is continuous in t , locally Lipschitz in A , and continuous in the remaining variables on bounded sets. Fix an initial condition $A(0) = A^0$. Then the classical Picard-Lindelöf theorem gives existence and uniqueness of a local solution on some interval $[0, T_{\max})$.

It remains to show continuous dependence on the initial condition, the mechanistic parameters, the correction parameters, and the patient descriptors. Let

$$p = (A^0, \theta, \phi, z)$$

collect these quantities, and let $A(t; p)$ denote the corresponding solution. Consider two parameter sets p_1 and p_2 , with associated solutions $A_1(t)$ and $A_2(t)$ defined on a common interval $[0, T]$ inside a bounded region where the local Lipschitz bound holds. Then

$$\frac{d}{dt}(A_1 - A_2) = H(A_1, p_1, u, t) - H(A_2, p_2, u, t).$$

Adding and subtracting $H(A_2, p_1, u, t)$ gives

$$\frac{d}{dt}(A_1 - A_2) = [H(A_1, p_1, u, t) - H(A_2, p_1, u, t)] + [H(A_2, p_1, u, t) - H(A_2, p_2, u, t)].$$

Taking norms and using local Lipschitz continuity in A , we obtain

$$\frac{d}{dt} \|A_1 - A_2\| \leq L \|A_1 - A_2\| + \omega(\|p_1 - p_2\|),$$

where L is a Lipschitz constant on the region and $\omega(\cdot)$ is a modulus of continuity with respect to the remaining variables. By Grönwall's inequality,

$$\|A_1(t) - A_2(t)\| \leq \left(\|A_1(0) - A_2(0)\| + \int_0^t \omega(\|p_1 - p_2\|) ds \right) e^{Lt}.$$

Hence

$$\sup_{0 \leq t \leq T} \|A_1(t) - A_2(t)\| \rightarrow 0 \quad \text{as } \|p_1 - p_2\| \rightarrow 0.$$

This proves continuous dependence on the initial condition, the parameters, and the patient descriptors. $\square$

Lemma A1. Let $A^0(t)$ be the solution of the mechanistic model (1), and let $A(t)$ be the solution of the hybrid model (2), with the same initial condition and the same input u . Assume that both trajectories remain in a bounded region Ω on a finite interval $[0, T]$. Assume also that

$$\|G(A, z, u, t; \phi)\| \leq \varepsilon$$

on Ω for some $\varepsilon > 0$. Then there exists a constant $K_T > 0$, depending on T and the Lipschitz bound of F on Ω , such that

$$\|A(t) - A^0(t)\| \leq K_T \varepsilon, \quad 0 \leq t \leq T.$$

Proof. Set

$$E_A(t) = A(t) - A^0(t).$$

Subtracting (1) from (2) gives

$$\frac{dE_A}{dt} = F(A, \theta, z, u, t) - F(A^0, \theta, z, u, t) + G(A, z, u, t; \phi).$$

Taking norms and using the Lipschitz continuity of F with respect to A , we obtain

$$\frac{d}{dt} \|E_A(t)\| \leq L \|E_A(t)\| + \varepsilon,$$

where L is a Lipschitz constant for F on Ω . Since the two systems have the same initial condition, we have

$$E_A(0) = 0.$$

Applying Grönwall's inequality yields

$$\|E_A(t)\| \leq \varepsilon \int_0^t e^{L(t-s)} ds = \varepsilon \frac{e^{Lt} - 1}{L}$$

if $L > 0$, and

$$\|E_A(t)\| \leq \varepsilon t$$

if $L = 0$. In either case, there is a constant $K_T > 0$ such that

$$\|A(t) - A^0(t)\| \leq K_T \varepsilon, \quad 0 \leq t \leq T.$$

This proves the lemma. $\square$

Proof of Proposition A2. Let

$$S_\theta^0(t) = \frac{\partial A^0(t)}{\partial \theta} \quad \text{and} \quad S_\theta(t) = \frac{\partial A(t)}{\partial \theta}$$

denote the sensitivity matrices of the mechanistic and hybrid models, respectively.

Differentiating (1) with respect to θ gives

$$\frac{d}{dt} S_\theta^0 = \frac{\partial F}{\partial A}(A^0, \theta, z, u, t) S_\theta^0 + \frac{\partial F}{\partial \theta}(A^0, \theta, z, u, t).$$

Differentiating (2) with respect to θ gives

$$\frac{d}{dt} S_\theta = \left(\frac{\partial F}{\partial A}(A, \theta, z, u, t) + \frac{\partial G}{\partial A}(A, z, u, t; \phi) \right) S_\theta + \frac{\partial F}{\partial \theta}(A, \theta, z, u, t) + \frac{\partial G}{\partial \theta}(A, z, u, t; \phi).$$

Define

$$E_S(t) = S_\theta(t) - S_\theta^0(t).$$

Subtracting the two sensitivity systems yields

$$\frac{dE_S}{dt} = \frac{\partial F}{\partial A}(A, \theta, z, u, t) E_S + R(t),$$

where

$$\begin{aligned} R(t) = & \left(\frac{\partial F}{\partial A}(A, \theta, z, u, t) - \frac{\partial F}{\partial A}(A^0, \theta, z, u, t) \right) S_\theta^0 + \frac{\partial G}{\partial A}(A, z, u, t; \phi) S_\theta \\ & + \left(\frac{\partial F}{\partial \theta}(A, \theta, z, u, t) - \frac{\partial F}{\partial \theta}(A^0, \theta, z, u, t) \right) + \frac{\partial G}{\partial \theta}(A, z, u, t; \phi). \end{aligned}$$

By smoothness of F on the bounded region, the differences

$$\frac{\partial F}{\partial A}(A, \theta, z, u, t) - \frac{\partial F}{\partial A}(A^0, \theta, z, u, t), \quad \frac{\partial F}{\partial \theta}(A, \theta, z, u, t) - \frac{\partial F}{\partial \theta}(A^0, \theta, z, u, t)$$

are bounded by a constant times $\|A(t) - A^0(t)\|$. By the assumptions of Proposition 2, the derivative $\partial_A G$, and also $\partial_\theta G$ when present, are bounded by ε on the region of interest. Since both trajectories remain in a bounded region on $[0, T]$, and the vector fields are smooth there, the corresponding sensitivity matrices also remain bounded on $[0, T]$, the corresponding sensitivity matrices remain bounded there as well. Using also Lemma A1, we therefore obtain

$$\|R(t)\| \leq c_1 \|A(t) - A^0(t)\| + c_2 \varepsilon \leq C\varepsilon$$

for some constant $C > 0$ on $[0, T]$.

Taking norms in the equation for E_S gives

$$\frac{d}{dt} \|E_S(t)\| \leq L \|E_S(t)\| + C\varepsilon,$$

where L bounds $\left\| \frac{\partial F}{\partial A} \right\|$ on the interval. Since

$$E_S(0) = 0,$$

Grönwall's inequality gives

$$\|E_S(t)\| \leq C\varepsilon \int_0^t e^{L(t-s)} ds \leq C_T \varepsilon, \quad 0 \leq t \leq T,$$

for a constant C_T depending on T and the bounds of the mechanistic system. This proves the result. $\square$

Proof of Proposition A3. Let J_0 be the sensitivity matrix of the parameter-to-observation map for the mechanistic model, restricted to the observation window of interest, and let J be the corresponding sensitivity matrix for the hybrid model. By assumption, J_0 has full column rank at the reference parameter value.

By Proposition 2, the sensitivity matrix of the hybrid state remains close to that of the mechanistic model. Since the observation map H is continuously differentiable on the region of interest, it follows that the corresponding observation sensitivity matrices also satisfy

$$\|J - J_0\| \leq C\varepsilon$$

for sufficiently small ε , where C is independent of ε on the observation window. Now the set of matrices with full column rank is open in the space of matrices. Since J_0 has full column rank, there exists $\delta > 0$ such that every matrix within distance δ of J_0 also has full column rank. Choose

$$\varepsilon_0 = \frac{\delta}{C}.$$

Then, whenever

$$\left\| \frac{\partial G}{\partial A} \right\| + \left\| \frac{\partial G}{\partial \theta} \right\| < \varepsilon_0,$$

it follows that

$$\|J - J_0\| < \delta,$$

and therefore J also has full column rank. Hence the hybrid model preserves local distinguishability of the mechanistic parameters on the same observation window.

This argument is local. It does not give a full structural identifiability result for the joint parameter set (θ, ϕ) . It shows only that the mechanistic parameters remain locally distinguishable under a sufficiently small constrained correction. $\square$

Appendix B. Patient-Level Paired Analyses

Paired differences were calculated as mechanistic MAE minus comparator MAE; positive values therefore indicate improvement over the mechanistic model. Confidence intervals were estimated using 10,000 subject-level bootstrap samples.

Table A1. Patient-level paired results for the polymyxin b held-out analysis.

Comparator	Median MAE Difference (95% CI)	Mean MAE Difference (95% CI)	Wilcoxon p	Rank-Biserial Effect	Improved, n/N (%)
Unconstrained hybrid	−0.120 (−0.263 to −0.020)	−0.324 (−0.651 to −0.069)	0.007	−0.483	13/40 (32.5%)
Constrained hybrid	−0.080 (−0.203 to 0.040)	−0.066 (−0.160 to 0.031)	0.106	−0.295	16/40 (40.0%)
Boosted-tree benchmark	−0.084 (−0.206 to 0.089)	−0.209 (−0.418 to −0.030)	0.146	−0.266	18/40 (45.0%)

Table A2. Patient-level paired results for the linezolid held-out analysis.

Comparator	Median MAE Difference (95% CI)	Mean MAE Difference (95% CI)	Wilcoxon <i>p</i>	Rank-Biserial Effect	Improved, <i>n</i> / <i>N</i> (%)
Unconstrained hybrid	−0.002 (−0.097 to 0.085)	−0.102 (−0.228 to 0.009)	0.401	−0.141	24/48 (50.0%)
Constrained hybrid	0.018 (−0.072 to 0.070)	−0.004 (−0.098 to 0.085)	0.867	0.029	26/48 (54.2%)
Boosted-tree benchmark	−0.008 (−0.047 to 0.091)	0.039 (−0.079 to 0.153)	0.467	0.122	23/48 (47.9%)

References

- Derendorf, H.; Schmidt, S. *Rowland and Tozer's Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications*, 5th ed.; Wolters Kluwer Health: Philadelphia, PA, USA, 2019.
- Darwich, A.S.; Polasek, T.M.; Aronson, J.K.; Ogungbenro, K.; Wright, D.F.B.; Achour, B.; Reny, J.-L.; Daali, Y.; Eiermann, B.; Cook, J.; et al. Model-Informed Precision Dosing: Background, Requirements, Validation, Implementation, and Forward Trajectory of Individualizing Drug Therapy. *Annu. Rev. Pharmacol. Toxicol.* **2021**, *61*, 225–245. <https://doi.org/10.1146/annurev-pharmtox-033020-113257>.
- Wicha, S.G.; Mårtson, A.-G.; Nielsen, E.I.; Koch, B.C.P.; Friberg, L.E.; Alffenaar, J.-W.C.; Minichmayr, I.K. From Therapeutic Drug Monitoring to Model-Informed Precision Dosing for Antibiotics. *Clin. Pharmacol. Ther.* **2021**, *109*, 928–941. <https://doi.org/10.1002/cpt.2202>.
- Minichmayr, I.K.; Dreesen, E.; Centanni, M.; Wang, Z.; Hoffert, Y.; Friberg, L.E.; Wicha, S.G. Model-Informed Precision Dosing: State of the Art and Future Perspectives. *Adv. Drug Deliv. Rev.* **2024**, *215*, 115421. <https://doi.org/10.1016/j.addr.2024.115421>.
- Kirubakaran, R.; Stocker, S.L.; Hennig, S.; Day, R.O.; Carland, J.E. Population Pharmacokinetic Models of Tacrolimus in Adult Transplant Recipients: A Systematic Review. *Clin. Pharmacokinet.* **2020**, *59*, 1357–1392. <https://doi.org/10.1007/s40262-020-00922-x>.
- Lu, J.; Deng, K.; Zhang, X.; Liu, G.; Guan, Y. Neural-ODE for Pharmacokinetics Modeling and Its Advantage to Alternative Machine Learning Models in Predicting New Dosing Regimens. *iScience* **2021**, *24*, 102804. <https://doi.org/10.1016/j.isci.2021.102804>.
- Valderrama, D.; Ponce-Bobadilla, A.V.; Mensing, S.; Fröhlich, H.; Stodtman, S. Integrating Machine Learning with Pharmacokinetic Models: Benefits of Scientific Machine Learning in Adding Neural Networks Components to Existing PK Models. *CPT Pharmacomet. Syst. Pharmacol.* **2024**, *13*, 41–53. <https://doi.org/10.1002/psp4.13054>.
- Hughes, J.H.; Keizer, R.J. A Hybrid Machine Learning/Pharmacokinetic Approach Outperforms Maximum a Posteriori Bayesian Estimation by Selectively Flattening Model Priors. *CPT Pharmacomet. Syst. Pharmacol.* **2021**, *10*, 1150–1160. <https://doi.org/10.1002/psp4.12684>.
- Antontsev, V.; Jagarapu, A.; Bunday, Y.; Hou, H.; Khotimchenko, M.; Walsh, J.; Varshney, J. A Hybrid Modeling Approach for Assessing Mechanistic Models of Small Molecule Partitioning In Vivo Using a Machine Learning-Integrated Modeling Platform. *Sci. Rep.* **2021**, *11*, 11143. <https://doi.org/10.1038/s41598-021-90637-1>.
- Staat, C.E.; Willis, C.; Taylor, P.J.; Tett, S.E. Population Pharmacokinetics of Tacrolimus in Adult Kidney Transplant Recipients. *Clin. Pharmacol. Ther.* **2002**, *72*, 660–669. <https://doi.org/10.1067/mcp.2002.129304>.
- Antignac, M.; Barrou, B.; Farinotti, R.; Lechat, P.; Urien, S. Population Pharmacokinetics and Bioavailability of Tacrolimus in Kidney Transplant Patients. *Br. J. Clin. Pharmacol.* **2007**, *64*, 750–757. <https://doi.org/10.1111/j.1365-2125.2007.02888.x>.
- Garcia-Prats, A.J.; Schaaf, H.S.; Draper, H.R.; Garcia-Cremades, M.; Winckler, J.; Wiesner, L.; Hesselring, A.C.; Savic, R.M. Pharmacokinetics, Optimal Dosing, and Safety of Linezolid in Children with Multidrug-Resistant Tuberculosis: Combined Data from Two Prospective Observational Studies. *PLoS Med.* **2019**, *16*, e1002789. <https://doi.org/10.1371/journal.pmed.1002789>.
- Hanafin, P.O.; Nation, R.L.; Scheetz, M.H.; Zavascki, A.P.; Sandri, A.M.; Kwa, A.L.; Cherng, B.P.Z.; Kubin, C.J.; Yin, M.T.; Wang, J.; et al. Assessing the Predictive Performance of Population Pharmacokinetic Models for Intravenous Polymyxin B in Critically Ill Patients. *CPT Pharmacomet. Syst. Pharmacol.* **2021**, *10*, 1525–1537. <https://doi.org/10.1002/psp4.12720>.
- Chen, N.; Guo, J.; Xie, J.; Xu, M.; Hao, X.; Ma, K.; Rao, Y. Population Pharmacokinetics of Polymyxin B: A Systematic Review. *Ann. Transl. Med.* **2022**, *10*, 231. <https://doi.org/10.21037/atm-22-236>.
- Al Lawati, H.; Al-Lawatia, M. A Structure-Preserving Hybrid Learning Framework for Compartmental Pharmacokinetic Models. *J. Pharmacokinet. Pharmacodyn.* **2026**, *53*, 22. <https://doi.org/10.1007/s10928-026-10031-0>.
- Hartman, P. Ordinary Differential Equations. In *Classics in Applied Mathematics*, 2nd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2002; Volume 38.

17. Quintairos, L.; Colom, H.; Millán, O.; Fortuna, V.; Espinosa, C.; Guirado, L.; Budde, K.; Sommerer, C.; Lizana, A.; López-Púa, Y.; et al. Early Prognostic Performance of miR155-5p Monitoring for the Risk of Rejection: Logistic Regression with a Population Pharmacokinetic Approach in Adult Kidney Transplant Patients. *PLoS ONE* **2021**, *16*, e0245880. <https://doi.org/10.1371/journal.pone.0245880>.
18. Liang, X.; Lai, G.; Yu, J.; Lin, T.; Wang, C.; Wang, W. Herbal Ingredient-Target Interaction Prediction via Multi-Modal Learning. *Inf. Sci.* **2025**, *711*, 122115. <https://doi.org/10.1016/j.ins.2025.122115>.
19. Jiang, M.; Cai, S.; Zhang, K.; Li, M.; Tian, Y.; Wang, T.; Alarcón Rodríguez, R.; Wang, Z.; Wu, S.; He, J.; et al. Deep Learning in Medical Image Analysis for Bone Tumor: A Comprehensive Survey. *Inf. Fusion* **2026**, *133*, 104305. <https://doi.org/10.1016/j.inffus.2026.104305>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.