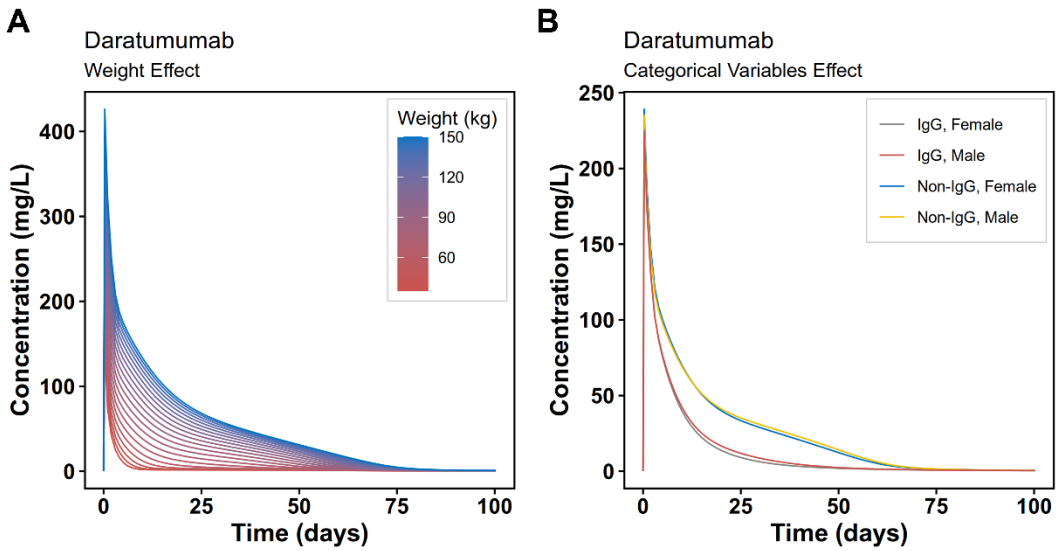
Sim2Real-PK framework for precision dosing: Integrating pharmacokinetic mechanisms and deep learning to decode clinical heterogeneity from sparse data

Lu-Yao Han[1#], Xiang Chen[1#], Huize Wang[1], Zhi-Long Zhang[2], Tian-Shuo Liu[2], Meng-Han Wu[1], Zi-Qi Wang[1], Wen-Bo Pang[1], Jia-Qian Li[1], Jin-Yi Wang[1], Xuan-De Kong[1], Man-Qing Jia[1], Haiping Hao[3*], Yang Yu[2*], Guo Yu[2*]

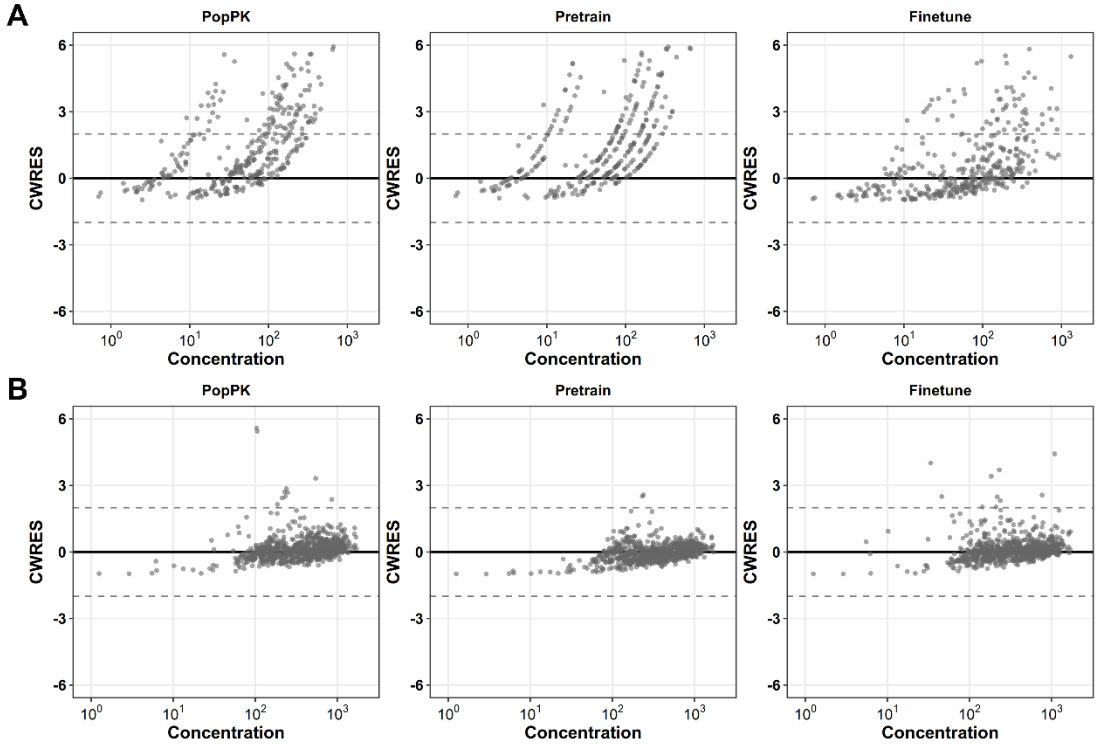
[1] School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 211198, PR China
[2] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, PR China
[3] State Key Laboratory of Natural Medicines, Jiangsu Provincial Key Laboratory of Drug Metabolism and Pharmacokinetics, China Pharmaceutical University, Nanjing, China

[#] Lu-Yao Han and Xiang Chen contributed equally to this work.
[*]Correspondence:
Guo Yu, PhD
School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, [#]24 Tongjia Lane, Nanjing, 210009 Jiangsu, China
Email: guoyu@cpu.edu.cn
Yang Yu, PhD
National Key Laboratory for Novel Software Technology, Nanjing University, No. 163 Xianlin Avenue, Qixia District, Nanjing, 210023 Jiangsu, China
Email: yuy@nju.edu.cn
Haiping Hao, PhD
State Key Laboratory of Natural Medicines, Jiangsu Provincial Key Laboratory of Drug Metabolism and Pharmacokinetics, China Pharmaceutical University, [#]24 Tongjia Lane, Nanjing, 210009 Jiangsu, China
Email: haipinghao@cpu.edu.cn

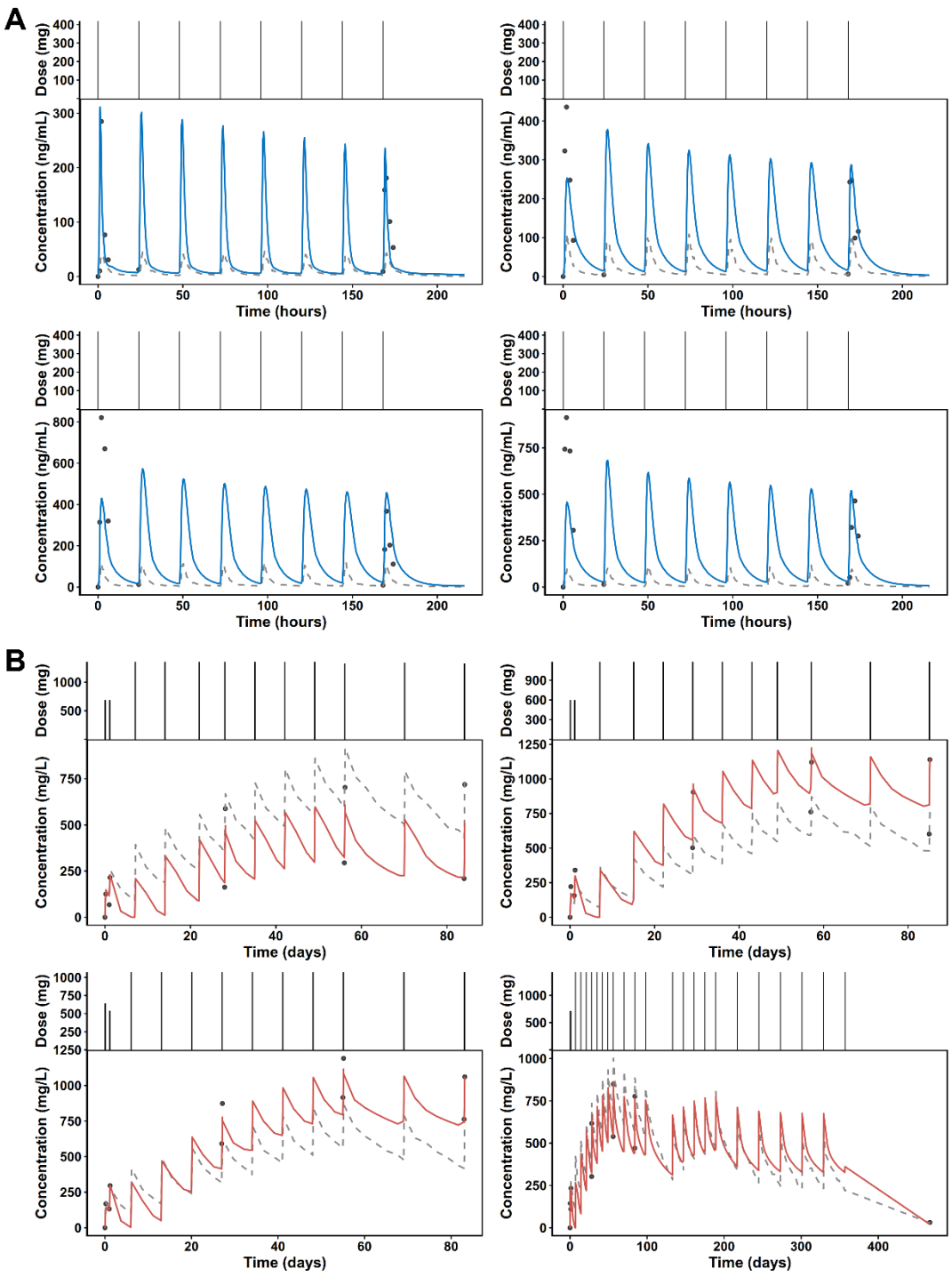
1. Supplementary Figures



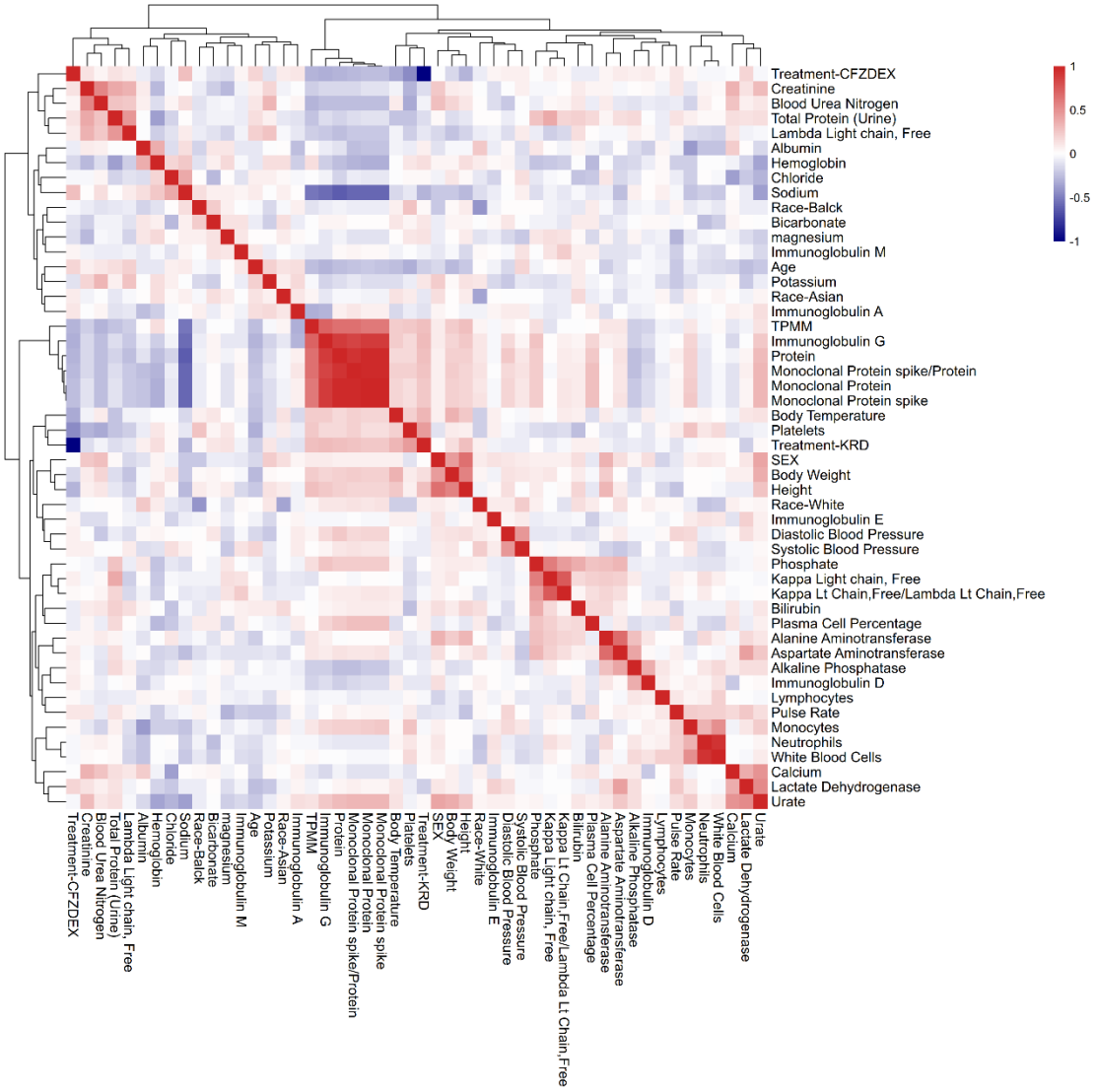
Supplementary Figure 1. Mechanistic sensitivity analysis of the pre-trained Sim2Real-PK model for Daratumumab. (A) Simulated single-dose concentration-time trajectories across a physiological body weight spectrum (60–150 kg), demonstrating the model's internalization of weight-dependent allometric scaling. (B) Comparative pharmacokinetic profiles stratified by key categorical covariates, isolating the distinct kinetic effects of sex (Male/Female) and immunoglobulin subtype (IgG vs. Non-IgG) on drug disposition.



Supplementary Figure 2. Diagnostic goodness-of-fit evaluation via Conditional Weighted Residuals (CWRES). Scatter plots display CWRES versus drug concentration (log scale) derived from the 5-fold cross-validation test sets for (A) Daratumumab and (B) Ibrutinib. Performance is stratified by modeling stage: the baseline Population PK benchmark (PopPK), the physics-informed Pre-trained initialization (Pretrain), and the final Fine-tuned Sim2Real-PK model (Finetune). The solid horizontal line indicates zero bias, while dashed lines demarcate the ± 2 interval.



Supplementary Figure 3. Representative individual pharmacokinetic profiles from cross-validation. Longitudinal concentration-time plots for randomly selected patients from the test set for (A) Ibrutinib (Blue lines) and (B) Daratumumab (Red lines). Black points denote observed clinical measurements. Solid colored curves represent the personalized predictions generated by the fine-tuned Sim2Real-PK model, while dashed grey curves represent the baseline Population PK benchmark. The vertical bars at the top of each panel indicate the timing and relative magnitude of dosing events.



Supplementary Figure 4. Correlation landscape of the high-dimensional clinical feature space. A hierarchically clustered heatmap visualizing the pairwise Pearson correlation coefficients across the full spectrum of patient covariates. The color gradient ranges from deep blue (strong negative correlation, -1) to dark red (strong positive correlation, +1). Dendrograms along the axes reveal natural clustering of physiologically related variables—such as the association between protein markers and immunoglobulin subtypes—highlighting the complex collinearity structure that is preserved during the virtual patient generation process.

2. Supplementary Table

Supplementary Table 1 The Feature groupings that correspond to the "leave-one-group-out" analysis

Feature Group	Description	Unit
Identified Covariates	Body Weight	kg
	Sex	-
	Albumin	g/mL
	Disease Type (IgG vs. Non-IgG)	-
Demographics	Age	years
	Height	cm
	Race: Asian	-
	Race: Black	-
	Race: White	-
Laboratory Tests	Alanine Aminotransferase	U/L
	Alkaline Phosphatase	U/L
	Aspartate Aminotransferase	U/L
	Lactate Dehydrogenase	U/L
	Total Bilirubin	mg/dL
	Creatinine	mg/dL
	Urate (Serum)	mg/dL
	Blood Urea Nitrogen	mg/dL
	Calcium	mg/dL
	Chloride	mmol/L
	Potassium	mmol/L
	Sodium	mmol/L
	Magnesium	mmol/L
	Phosphate	mg/dL
	Bicarbonate	mmol/L
	Neutrophils	cells/μL
	Monocytes	cells/μL
	Hemoglobin	g/dL
	Platelets	cells/μL
	Lymphocytes	cells/μL
	White Blood Cells	cells/μL
Vital Signs	Diastolic Blood Pressure	mmHg
	Systolic Blood Pressure	mmHg
	Body Temperature	°C
	Pulse Rate	beats/min
Immunoglobulins	Immunoglobulin G	g/L
	Immunoglobulin A	g/L
	Immunoglobulin M	g/L
	Immunoglobulin D	g/L
	Immunoglobulin E	IU/mL

Feature Group	Description	Unit
M Protein	Monoclonal Protein (Serum)	g/dL
	Monoclonal Protein Spike	g/dL
	Total Protein	g/dL
	M-Protein Spike / Total Protein Ratio	-
	Total Protein (Urine)	g/24h
Free Light Chains and Plasma cell percentage	Plasma Cell Percentage (Bone Marrow)	%
	Kappa Free Light Chain	mg/dL
	Lambda Free Light Chain	mg/dL
	Kappa/Lambda Free Light Chain Ratio	-
Therapies	Carfilzomib + Dexamethasone	-
	Carfilzomib + Lenalidomide + Dexamethasone	-

Supplementary Table 2 Hyperparameters of Pretraining and Finetuning

Hyperparameter	**Pre-training**	**Fine-tuning**
Model architecture		
Dynamic model	hidden_size: 768, layers: 4,	
Identified Covariates Encoder	activation: SiLU, layers: 3, hidden_size: 768, output: 768, dropout: 0.1	
Action Encoder	activation: SiLU, layers: 3, hidden_size: 768, output: 768, dropout: 0.1	
Observation Encoder/Decoder	activation: SiLU, layers: 3, hidden_size: 768, output: 768, dropout: 0.1	
Time Encoder	activation: ReLU, layers: 3, hidden_size: 64, output: 16	
Extended Covariates Encoder	/	activation: ReLU, layers: 3, hidden_size: 64
Residual Decoder	/	activation: ReLU, layers: 5, hidden_size: 128/64
Training strategy		
Training mode	two-stage-adaptive	linear
Trainable modules	— (full model)	Identified Covariates Encoder Extended Covariates Encoder Residual Decoder
Update observation decoder	Yes	No
Loss scale		
Reconstruction Loss	1.0	1.0
Consistency Loss	1.0	1.0
AUC Loss	0.01	0
C_{max} Loss	0.8	0.02
C_{min} Loss	1.0	0.02
Rollout Loss	1.2	0
Optimization		
Segment length	64	128
Warmup length	10	10
Total epochs	100	300
Batch size	64	64
Learning rate	1×10^{-4}	5×10^{-4}

3. Supplementary Methods

3.1 Synthetic Data Generation via ODEs

Ground-truth synthetic datasets were generated using standard Ordinary Differential Equations (ODEs) to represent the pharmacokinetic mechanisms of the target drugs. Synthetic pre-training data for both drugs were produced in two steps: (1) dose–covariates generation, which defined virtual subjects with dosing schedules and covariates; (2) PK simulation, which read these schedules, solved the corresponding ODEs per subject, and wrote concentration–time series. For both drugs, the design combined variable-dose and constant-dose regimens to cover diverse dosing patterns. Covariates were sampled from all combinations of discrete levels (e.g., WEIGHT × ALB × TPMM × SEX for daratumumab; meal × CYP3Ai for ibrutinib). The dense ODE output was downsampled using an adaptive key-point scheme that retains peaks, troughs, dosing times, and gradient-based spacing so that training data preserve curve shape while reducing size.

Ibrutinib was modeled with a two-compartment model with first-order absorption and linear elimination. The state variables are depot (A1), central (A2), and peripheral (A3); absorption into the central compartment is first-order (rate constant k_a), and elimination and distribution are linear. The typical values of parameters (e.g., CL/F, Vc/F, Q/F, Vp/F, k_a) were taken from published PopPK literature, with meal (fasted vs fed) and CYP3Ai (no inhibitor, moderate inhibitor, and strong inhibitor) affecting zero-order input duration (D1) and bioavailability (F1). Doses were fixed per administration (e.g., 140, 280, 420 mg, reflecting tablet strengths), with a constant inter-dose interval per subject chosen at random from 24, 48, 72, or 96 hours over a 240-hours simulation. There is no weight-based scaling or phase-based interval escalation; this reflects oral, once-daily–like or less frequent dosing.

Daratumumab was modeled with a two-compartment model with target-mediated drug disposition (TMDD). The ODEs include linear (non-specific) clearance (CL) and saturable (target-mediated) clearance ($V_{\max}(t)C/(K_M + C)$), with time-varying $V_{\max}(t) = V_{\max,0} * \exp(-k_{\text{des}}t)$ to capture declining target-mediated clearance over time. Central (A1) and peripheral (A2) compartments are linked by linear distribution. PopPK parameters (e.g., CL, V1, Q, V2, K_M, $V_{\max,0}$, k_{des}) and their covariate relationships (weight, albumin, type of multiple myeloma, sex) were taken from a PopPK model (e.g., allometric scaling for CL and V1, albumin and type of multiple myeloma on CL, sex on V1). Doses were weight-based (e.g., 8–20 mg/kg), drawn from

discrete mg/kg options. To reflect label and real-world practice, the dosing schedule was phase-based: an initial period (e.g., before ~day 49) used shorter intervals (~6–9 days, Q1W-like); a middle period (~day 49–167) used ~13–16 days (Q2W); a later period (after ~day 167) used ~20–24 days (Q3W). Inter-dose intervals were therefore time-varying within each subject. For high first doses (e.g., ≥16 mg/kg), the first dose could be split over two consecutive days (e.g., 80% probability) to mimic mitigation of infusion-related reactions. Infusion duration was explicitly modeled: first dose median ~7 h, second ~4.3 h, subsequent ~3.4 h (with ±20% variation), implemented as piecewise-constant infusion rates in the ODEs. Total simulation length was 1000 days to cover long-term maintenance.

3.2 Training Strategy Comparison

To validate the Two-Stage Adaptive Curriculum, we compared it to an Epoch-Based Linear Schedule baseline. Both strategies control (i) scheduled sampling (teacher forcing vs. autoregressive inputs) and (ii) rollout loss (loss on free-run predictions). They differ in how the teacher-forcing schedule and the rollout-loss weight are chosen.

For both strategies, the world model is trained to predict the next observation from the previous latent state, actions, population features, and time deltas. At each step t, the input to the RNN is either the ground-truth observation (teacher forcing) or the model's previous prediction (autoregressive). The probability of using ground-truth is set by a scalar prob per epoch: at step t, ground-truth is used with probability 1−prob (and with probability 1 at t=0).

For the Epoch-Based Linear Schedule baseline, the probability of using ground-truth inputs was decayed linearly over epochs, independent of performance. In the implementation, prob=epoch/E, where E is the total number of training epochs (e.g. E=100 in the comparison). Thus, the teacher-forcing probability is 1−epoch/E, i.e. it decreases linearly from 1.0 to 0.0 over the first E epochs. The weight on the autoregressive rollout loss was fixed at 1.0 for the whole run; no adaptation to reconstruction or rollout performance was applied.

For the Two-Stage Adaptive Curriculum, the schedule is two-phase rather than linear. For the first half of training (epoch<E/2), prob increases from 0 to 0.8; for the second half, it increases from 0.8 to 1.0. Thus, teacher forcing is reduced more gradually and never drops to zero in the first phase, giving the model more time to adapt before full autoregression. The weight on the autoregressive rollout loss (w_{adapt}) is performance-gated. A baseline reconstruction loss is estimated as the mean

over the first few epochs. For the first half of training, the rollout loss weight is set to zero, so no gradient is applied from free-run rollout in this phase. For the second half, the weight is computed from the ratio of the current reconstruction and consistency loss to the baseline. In the implementation, the ratio is combined via a geometric mean and scaled. Thus, w_{adapt} is dynamic: when the model fits the data well (current loss below baseline), the rollout term is down-weighted to limit instability and forgetting; when loss is still high, rollout receives full weight.

3.3 Mechanistic Verification Protocols

To verify scaling laws of dose-exposure relationship, simulations were run for a reference virtual patient under a range of single doses (e.g., 140 mg to 420 mg for Ibrutinib; 4 mg/kg to 20 mg/kg for Daratumumab). The resulting $AUC_{0-\infty}$ was regressed against dose to assess linearity (R^2).

To verify the identified covariates sensitivity of the pretrained model, simulations were conducted by varying a single identified covariate (e.g., Albumin from 20 to 50 g/L) while holding all other parameters constant. The resulting $AUC_{0-\infty}$ were visualized to confirm alignment with known allometric scaling principles.

3.4 Real-World 5-Fold Cross-Validation

The clinical dataset was partitioned into 5 folds at the subject level (ensuring all timepoints from one patient belong to the same fold). For each fold iteration: The pretrained model was initialized using the synthetic cohort relevant to the drug type. In the finetuning stage, the "Freeze-Thaw" strategy was applied using the training set (4 folds). The externded high-dimensional Encoder and Residual Decoder were initialized with zero weights. Predictions were generated for the held-out test set (the left fold) in a fully autoregressive mode (free-run). Metrics (RMSE, F20/F50) were computed and averaged across all 5 folds.

3.5 Feature Ablation and Grouping

Feature importance was assessed using a "Leave-One-Group-Out" approach. Clinical features were categorized into functional groups (Detailed in Supplementary Table 1): Demographics, Vital Signs, Lab Tests, Immunoglobulins, M-Proteins, Free Light Chains, and Concomitant Therapies. For each experiment, one specific group was masked (values set to zero) during both training and inference, while maintaining all other architecture components. The relative increase in RMSE compared to the "Full Model" quantified the group's contribution.

3.6 Action Search for Dose Optimization

To demonstrate the framework's utility in personalized medicine, we implemented a discrete optimization algorithm to tailor dosing schedules. We defined the dosing strategy as a sequence of varying frequencies—weekly (n_{QW}), bi-weekly (n_{Q2W}), and tri-weekly (n_{Q3W})—while fixing the total number of doses to match standard clinical duration. The optimization goal was to identify the optimal allocation of doses across these phases for a specific patient, effectively balancing the duration of induction and consolidation periods against individual metabolic profiles.

The search was guided by a penalty function designed to minimize violations of a therapeutic window derived from established exposure-response analyses. On the efficacy side, the minimum trough concentration was set at $236\ \mu g/mL$, a threshold required to achieve 99% target saturation (EC_{99}^{TAR}) and induce maximal clinical effect. Conversely, the safety threshold constrained the maximum peak concentration below $907\ \mu g/mL$. This upper limit corresponds to the boundary of the highest exposure quartile, where clinical data indicates a numerical increase in the rate of adverse events, particularly infections.

To navigate this discrete state space, we employed a greedy neighborhood search strategy. Starting from a standard protocol, the algorithm iteratively explores "neighboring" strategies by slightly adjusting the interval of the dose. For each candidate strategy, the fine-tuned Sim2Real-PK model predicts the full pharmacokinetic trajectory to evaluate compliance with the defined constraints. The algorithm accepts adjustments that reduce the penalty—thereby reducing toxicity risk while maintaining target saturation—and terminates when no further improvements can be found in the local neighborhood.

3.7 Virtual Twin Simulation for Subtype Analysis

To investigate PK heterogeneity in non-IgG patients (where specific subtype labels are often missing in PopPK models), a generative model was built. A covariance matrix σ was computed from the extended high-dimensional clinical features of the training population. Virtual twins were generated by sampling from the multivariate normal distribution $\mathcal{N}\ \mu, \sigma$, conditioned on fixed identified covariates. Two virtual cohorts were synthesized. an IgA-Dominant group (high IgA Z-scores) and a light chain-dominant group (high light chain Z-scores), with identical weight and albumin. Their PK trajectories were simulated to reveal divergences driven solely by extended high-dimensional features.

4. Supplementary Algorithms

4.1 Supplementary Algorithm 1

Algorithm 1: Pre-training with Two-Stage Adaptive Strategy

Input: Source Dataset $\mathcal{D}_{src}$, Total Epochs E_{max}
Output: Pre-trained Parameters θ

Initialize θ randomly; Baselines $\bar{\mathcal{L}}_{rec} \leftarrow \emptyset, \bar{\mathcal{L}}_{cons} \leftarrow \emptyset$
for *epoch* $e \leftarrow 1$ **to** E_{max} **do**
  `// === Two-Stage Adaptive Curriculum ===`
  if $e < E_{trans}$ *(Phase I)* **then**
    $P_{autoreg} \leftarrow (e/E_{trans}) \cdot 0.8$ `// Encourage dependency`
    $w_{adapt} \leftarrow 0.0$ `// Disable PK constraints`
  end
  else
    $P_{autoreg} \leftarrow 0.8 + \frac{e-E_{trans}}{E_{max}-E_{trans}} \cdot 0.2$ `// Towards autonomy`
    Update baselines $\bar{\mathcal{L}}_{rec}, \bar{\mathcal{L}}_{cons}$ from history
    $w_{adapt} \leftarrow$ Active `// Enable constraints`
  end
  for *batch* $(\mathbf{y}, \mathbf{a}, \mathbf{x}, \Delta\mathbf{t}) \sim \mathcal{D}_{src}$ **do**
    `// 1. Standard Training Step`
    $\hat{\mathbf{y}}, \mathbf{z} \leftarrow \text{Forward}(\mathbf{y}, \mathbf{a}, \mathbf{x}, \Delta\mathbf{t}, P_{autoreg})$
    $\mathcal{L}_{MSE} \leftarrow \mathcal{L}_{rec}(\hat{\mathbf{y}}, \mathbf{y}) + \mathcal{L}_{cons}(\mathbf{z}, \mathbf{y})$
    `// 2. Adaptive Weight Calculation`
    if w_{adapt} *is Active* **then**
      $r_{rec} \leftarrow \mathcal{L}_{rec}/\bar{\mathcal{L}}_{rec}; \quad r_{cons} \leftarrow \mathcal{L}_{cons}/\bar{\mathcal{L}}_{cons}$
      $ratio \leftarrow \sqrt{r_{rec} \cdot r_{cons}}$ `// Geometric Mean`
      $w_{adapt} \leftarrow \text{clip}(1.5 \cdot ratio, 0.1, 1.0)$
    end
    `// 3. Segment Rollout for PK Metrics`
    Sample segment $[t_{start}, t_{end}]$; Split into Warm-up & Free-run
    Initialize state with Warm-up (Teacher Forcing)
    Generate trajectory $\hat{\mathbf{y}}_{free}$ via Free-run (Autoregression)
    `// Compute Pharmacological Penalties`
    $\mathcal{L}_{rollout} \leftarrow \text{MSE}(\hat{\mathbf{y}}_{free}, \mathbf{y}_{free})$
    $\mathcal{L}_{PK} \leftarrow \lambda_2\mathcal{L}_{AUC} + \lambda_3\mathcal{L}_{peak} + \lambda_4\mathcal{L}_{trough}$
    `// 4. Compound Loss Optimization (Eq. 9)`
    $\mathcal{L}_{total} \leftarrow \mathcal{L}_{MSE} + w_{adapt} \cdot (\lambda_1\mathcal{L}_{rollout} + \mathcal{L}_{PK})$
    Update $\theta \leftarrow \text{Optimizer}(\nabla\mathcal{L}_{total})$
  end
end

4.2 Supplementary Algorithm 2

Algorithm 2: Fine-tuning with Extended Residual Correction

Input: Target Dataset $\mathcal{D}_{tgt}$, Pre-trained θ_{base}
Output: Fine-tuned Parameters $\{\theta_{base}, \theta_{res}\}$
// 1. Initialization Zero-Start Trick
Load θ_{base} (Freeze or Low Learning Rate)
Initialize $\theta_{res} = \{\text{Enc}_{high}, \text{Dec}_{res}\}$
Zero-Initialize Dec_{res} output layer so $\delta(\cdot) \approx 0$ initially
while *not converged* **do**
  Sample batch $(\mathbf{y}, \mathbf{a}, \mathbf{x}, \Delta\mathbf{t}) \sim \mathcal{D}_{tgt}$
  // Feature Split: Identified vs. Extended
  Split $\mathbf{p} \to [\mathbf{x}_{ide}, \mathbf{x}_{ext}]$
  for *step t in Sequence* **do**
    $\Delta\mathbf{t} \leftarrow \mathbf{t}_t - \mathbf{t}_{t-1}$
    // A. Base Path (General Dynamics)
    $\mathbf{z}_t \leftarrow \text{Dynamic}_{\theta_{base}}(\ldots, \mathbf{p}_{base}, \Delta\mathbf{t})$
    $\hat{\mathbf{y}}_{base} \leftarrow \text{Dec}_{obs}(\mathbf{z}_t)$
    // B. Residual Path (Patient Specificity)
    if $\mathbf{x}_{ext}$ *is available* **then**
      $\mathbf{e}_{ext} \leftarrow \text{Enc}_{ext}(\mathbf{x}_{ext})$ // Encode Extended Feature
      $\mathbf{r}_{in} \leftarrow \text{Concat}(\mathbf{z}_t, \mathbf{e}_{ext}, \text{Enc}_{time}(\Delta\mathbf{t}))$
      $\boldsymbol{\delta}_t \leftarrow \text{Dec}_{res}(\mathbf{r}_{in})$ // Compute Correction
      $\hat{\mathbf{y}}_{final} \leftarrow \text{ReLU}(\hat{\mathbf{y}}_{base} + \boldsymbol{\delta}_t)$
    end
    else
      $\hat{\mathbf{y}}_{final} \leftarrow \hat{\mathbf{y}}_{base}$
    end
  end
  $\mathcal{L}_{rollout} \leftarrow \text{MSE}(\hat{\mathbf{y}}_{free}, \mathbf{y}_{free})$
  $\mathcal{L}_{PK} \leftarrow \lambda_2\mathcal{L}_{AUC} + \lambda_3\mathcal{L}_{peak} + \lambda_4\mathcal{L}_{trough}$
  $\mathcal{L}_{ft} \leftarrow \mathcal{L}_{MSE} + w_{adapt} \cdot (\lambda_1\mathcal{L}_{rollout} + \mathcal{L}_{PK})$
  Update $\{\theta_{base}, \theta_{res}\} \leftarrow \text{Optimizer}(\nabla\mathcal{L}_{ft})$
end