PAGE 2026 | Population Approach Group Europe

CVAE–LASSO: Integrated L1-Regularized Covariate Selection within a Conditional Latent Framework for Population Pharmacokinetic Modelling

Diego Perazzolo[1], **Enrico Grisan**[2]

[1] *University of Padova, Dept. of Cardiac, Thoracic, Vascular Sciences and Public Health, Padova, Italy*
diego.perazzolo.1@studenti.unipd.it

[2] *London South Bank University, School of Engineering, London, UK*
enrico.grisan@lsbu.ac.uk

Background & Objectives

Population PK (PopPK) modelling quantifies inter-individual variability and identifies covariates that drive drug exposure, enabling individualised dosing. Traditional nonlinear mixed-effects models with stepwise covariate search are slow and struggle with nonlinear effects.

Our earlier VAE–LASSO (PAGE 2025) kept representation learning and sparse inference separate; here we propose an **integrated** CVAE–LASSO where sparse covariate selection directly conditions PK-profile generation.

Aims

- Embed an **L1-regularised LASSO** module inside the conditional VAE so covariates condition both encoder and decoder during PK-profile reconstruction.
- Verify that the sparse coefficients recover the truly influential covariates from only 300 training profiles, with two noise covariates as a sanity check.
- Deliver a fast, interpretable covariate-screening tool for PopPK analyses.

Methods

Case study & data generation

Synthetic dorzagliatin PK profiles were simulated from a **two-compartment model** with sequential zero- and first-order absorption and first-order elimination, based on a published PopPK model. Inter-compartmental clearance (Q/F), peripheral volume (Vp/F), and absorption rate constant (Ka) were fixed to their reference values. Covariate effects on CL/F, Vc/F, and absorption duration (D_1) were described as reported:

$$CL_i = 10.4 \times \exp(0.255 \times \ln(TBW_i/69) - 0.103 \times \ln(AST_i/18) - 0.135 \times \ln(AGE_i/55))$$

$$V_{c,i} = 80.6 \times \exp(0.553 \times \ln(TBW_i/69) - 0.170 \times SEX_i)$$

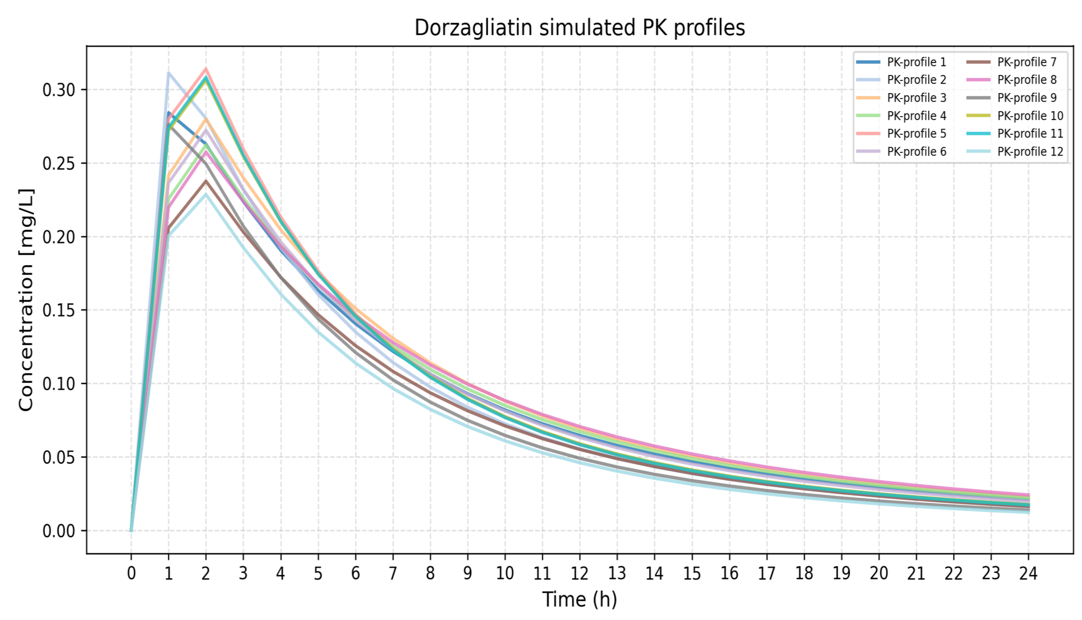
$$D_{1,i} = 0.418 \times \exp(0.816 \times FOOD_i)$$

15 covariates (AGE, TBW, HT, BMI, RBC, ALB, ALT, AST, GGT, CR, TBIL, SEX, FOOD) were drawn from truncated-normal distributions using reported medians and bounds;

2 noise covariates (extra_1, extra_2) were added to probe specificity.

2 datasets: 300 training profiles (realistic clinical size) and **2000 independent test** profiles 24 h window after a single 25 mg oral dose.

Covariate	Median	Min–Max
AGE (yr)	55.0	19–74
TBW (kg)	69.0	40–110
HT (cm)	165	141–186
BMI	25.2	18.2–34.9
RBC (10^{12}/L)	4.85	2.81–6.92
ALB (g/L)	46.2	33.2–57.0
ALT (U/L)	19.0	2.5–110
AST (U/L)	18.0	8–74
GGT (U/L)	24.0	7–334
CR (μmol/L)	67.0	35–943
TBIL (μmol/L)	10.0	1.25–40

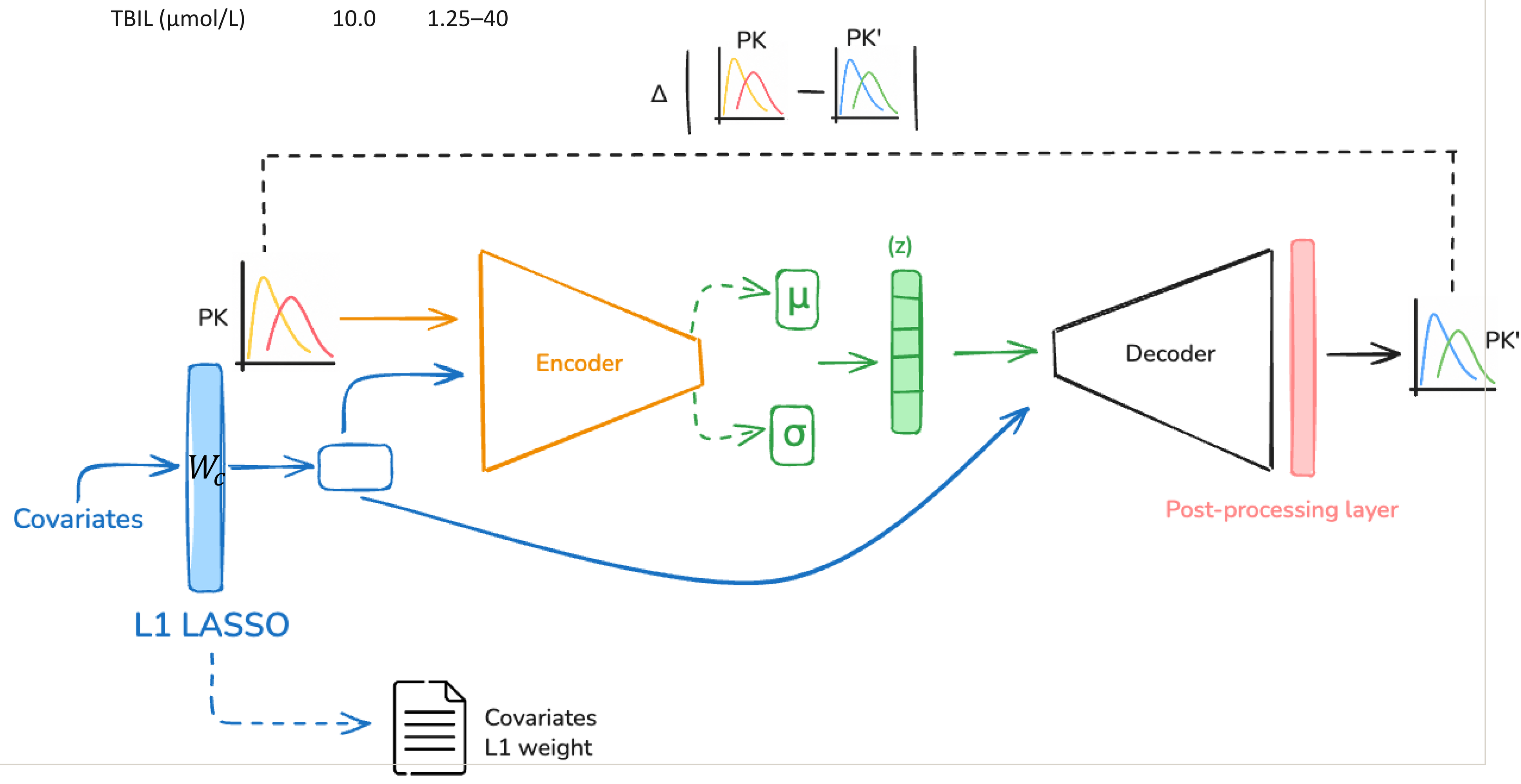


CVAE–LASSO architecture

Covariates pass through a **LASSO layer** enforcing sparsity via an L1 penalty; the selected sparse representation conditions **both encoder and decoder** of the CVAE. Continuous covariates were min–max scaled; categorical variables one-hot encoded.

$$\mathcal{L}(\theta) = \frac{1}{N}\Sigma_i \|PK(t)_i - PK'(t)_i\|_1 + \beta\, D_{KL}(q_\varphi(z|PK(t),c) \,\|\, p(z)) + \lambda \|W_c\|_1$$

λ tested over {0.01, 0.1, 0.25, 0.5, 0.75, 1, 2, 3}. Reconstruction assessed with MAE, MAPE and dynamic time-warping (DTW) distance. A quadratic-interpolation **post-processing layer** refined smoothness and physiological plausibility.



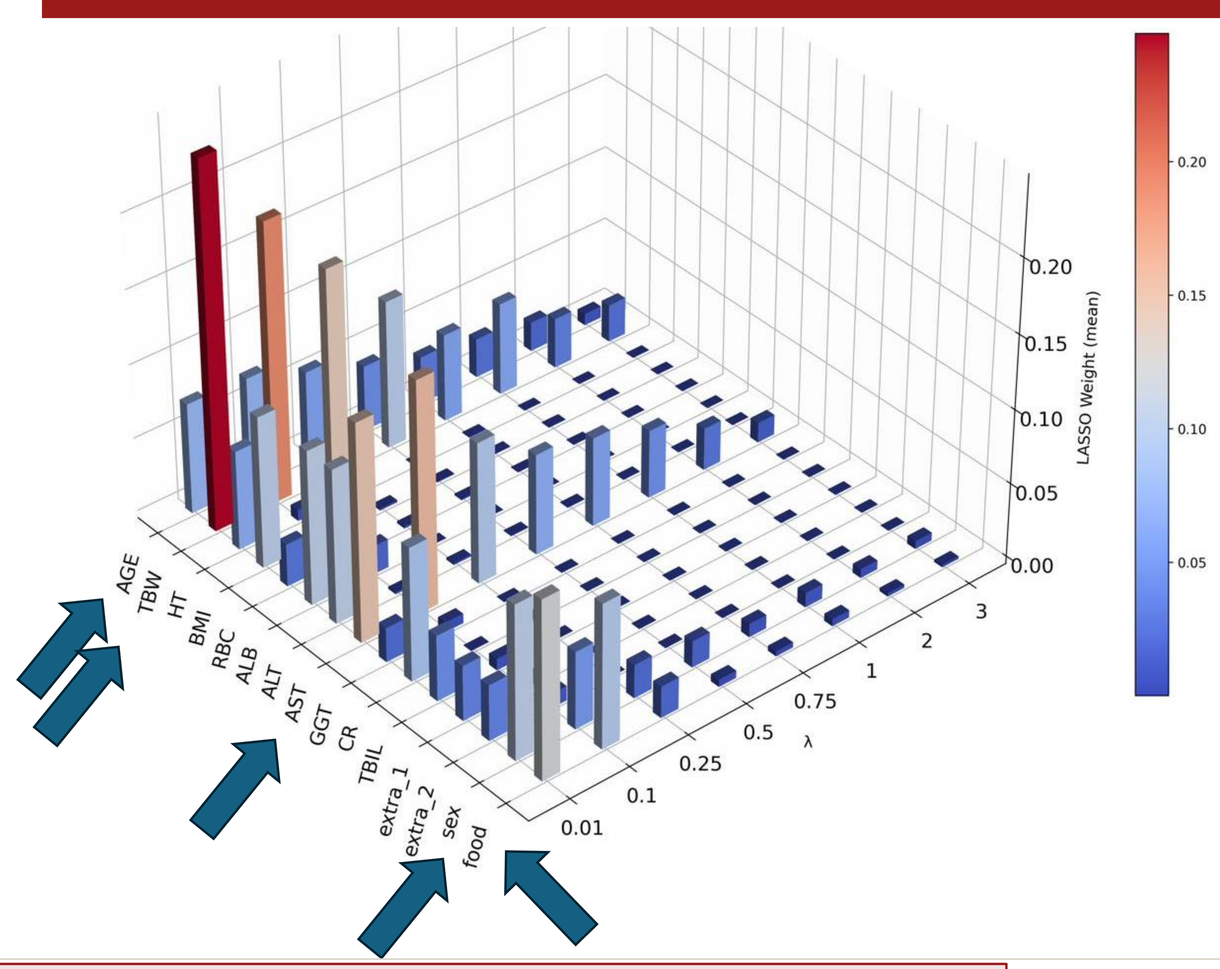
Results: reconstruction of PK profiles

λ	MAE (mg/L)	MAPE (%)	DTW (–)
0.01	0.0016	1.993	0.012
0.1	**0.0015**	**1.984**	**0.011**
0.25	0.0086	4.782	0.138
0.5	0.0038	2.808	0.066
0.75	0.0032	2.797	0.041
1.0	0.0022	2.271	0.027
2.0	0.0028	2.813	0.031
3.0	0.0024	2.581	0.031

Table 1. Reconstruction metrics (MAE, MAPE, DTW) across L1 penalty λ. Best performance at λ = 0.1.

Accurate reconstructions across all λ settings: MAE in the 10^{-3} mg/L range and MAPE consistently below 5%, with DTW close to zero indicating strong temporal alignment between generated and reference trajectories.

Results: covariate discovery



Mean LASSO weights across covariates and L1 penalty λ. Tall, warm-coloured bars (AGE, TBW, SEX, FOOD) mark dominant covariates; weights shrink progressively as λ increases, giving tunable sparsity.

Covariates consistently recovered

Across all λ values, AGE, TBW, SEX, and FOOD consistently retained higher weights, confirming their predominant role in PK profile reconstruction. Weaker covariates and the injected noise variables (extra_1, extra_2) are driven to zero and excluded at higher λ.

Conclusions

- Embedding LASSO as a conditioning mechanism within a CVAE yields an integrated, interpretable framework that jointly reconstructs PopPK profiles and produces sparse, data-driven covariate selection.
- Trained on only 300 profiles and evaluated on 2000 independent profiles, it kept stable reconstruction quality and recovered the most influential covariates in a more complex PK structure than our previous work.
- λ gives practical control over sparsity, enabling rapid covariate prioritization and reducing reliance on exhaustive stepwise searches.

Limitations & next steps Current results use simulated, regularly sampled data. Future work will extend to irregular sampling, missing observations and multi-centre clinical datasets, benchmarking against standard PopPK frameworks.

LSBU Biosciences and Bioengineering
University Research Centre