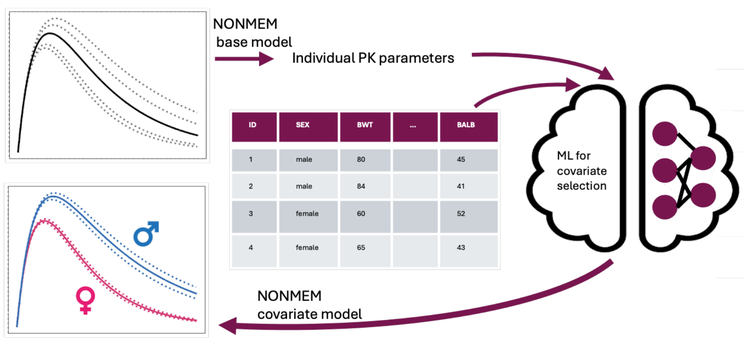
Stochastic Gate Neural Networks for Automatic Covariate Selection in Pharmacometrics Population Modeling

Marija Kekic[1], Oleg Stepanov[2], Andrzej Nowojewski[3], Sam Richardson[3], Itziar Irurzun Arana[2], Jacob Leander[4], Diansong Zhou[5], Weifeng Tang[6], Richard Dearden[3], Megan Gibbs[6]

[1]Imaging & Data Analytics, Clinical Pharmacology & Safety Sciences, R&D BioPharmaceuticals, AstraZeneca, Barcelona, Spain
[2]Clinical Pharmacology & Quantitative Pharmacology, Clinical Pharmacology & Safety Sciences, R&D BioPharmaceuticals, AstraZeneca, Cambridge, UK
[3]Imaging & Data Analytics, Clinical Pharmacology & Safety Sciences, R&D BioPharmaceuticals, AstraZeneca, Cambridge, UK
[4]Clinical Pharmacology & Quantitative Pharmacology, Clinical Pharmacology & Safety Sciences, R&D BioPharmaceuticals, AstraZeneca, Gothenburg, Sweden
[5]Clinical Pharmacology & Quantitative Pharmacology, Clinical Pharmacology & Safety Sciences, R&D BioPharmaceuticals, AstraZeneca, Waltham, MA, USA
[6]Clinical Pharmacology & Quantitative Pharmacology, Clinical Pharmacology & Safety Sciences, R&D BioPharmaceuticals, AstraZeneca, Gaithersburg, MD, USA

Abstract



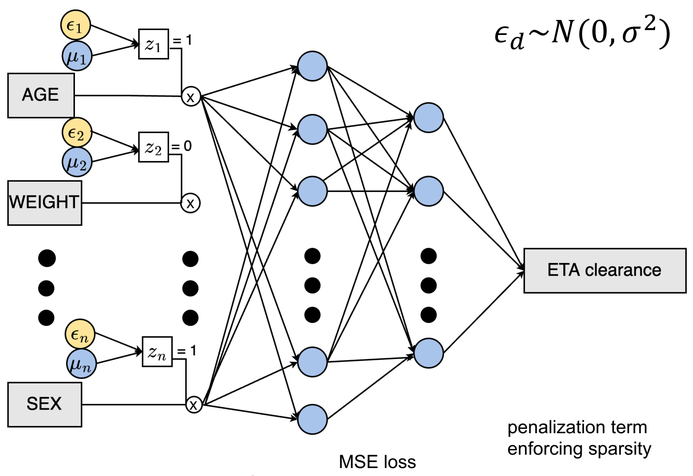
Introduction

Population pharmacokinetic (PK) models describe the behavior of drugs in the body and are usually constructed within a nonlinear mixed-effects framework. The modeling process typically unfolds in two steps: first, a structural model is developed, where type of absorption, clearance, or the number of compartments are chosen. The second stage involves searching for covariates that stratify the population and clearly affect drug behavior.
Covariates are chosen based on clinical and statistical relevance - typically determined by running a time-consuming stepwise covariate selection algorithm [1].
Recently, there have been attempts to employ fast Machine Learning (ML) methods to pre-select relevant covariates by searching for patterns in estimated individual PK parameters [2, 3].
The objective of this study is to explore Neural Networks (NN) with Stochastic Gates [4] layer, which provide a one-step training and feature selection algorithm.

Methods

We fit a base, covariate-free model to obtain estimates of individual random effects (ETAs) for which we use NONMEM software. Instead of using empirical Bayes estimates for each individual ETA, we use samples from conditional distributions [5].
A NN with Stochastic Gate Layer implemented in PyTorch is used to predict individual ETA values from patient covariates. The layer introduces an additional hyperparameter (lambda) controlling the penalization for the number of input covariates.

$$z_d = \max(0, \min(1, \mu_d + \epsilon_d))$$
$$\epsilon_d \sim N(0, \sigma^2)$$



$$Loss = \frac{1}{N}\sum_{n}\left(Y_{true} - Y_{pred}\right)^2 + \lambda||Z||_0$$

Two terms must be of the same order of magnitude to ensure effective learning!

Final gates are a mean value of gates probability obtained from a 5-fold cross validation.
All hyperparameters (given below), including NN architecture, learning parameters and lambda are manually tuned on one synthetic dataset and kept the same for the rest of the datasets.
Running time for one PK parameter is 15-20 mins on a single GPU.

Number of layers	3	Weight decay	0.001
Nodes per layer	50	Activation	tanh
Dropout	0.5	Optimizer	Adam
lambda	0.2	Learning rate	0.001

Results: validation on simulated data

- To investigate the limits of performance, we tested the algorithm on simulated models, where the base model is a single dose 2-compartment model with first order absorption and elimination.
- Categorical (CAT) and continuous (CON) covariate effects are added to clearance (CL) parameter as:

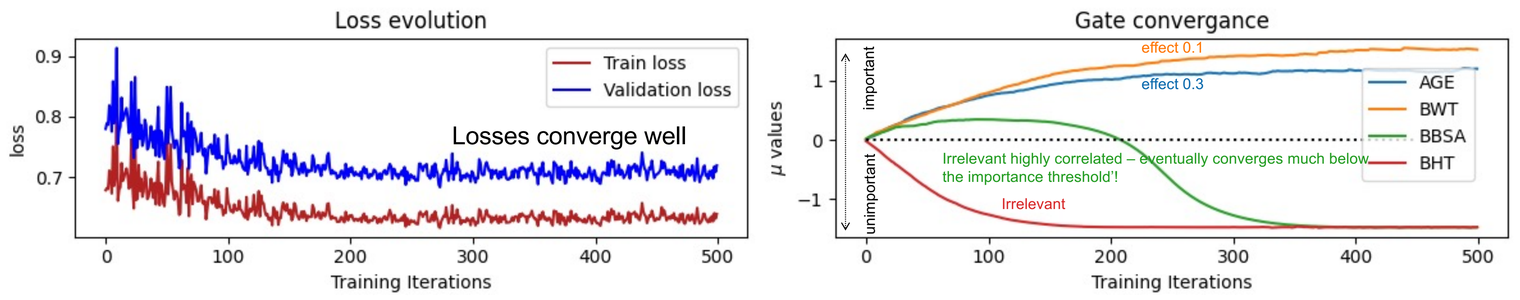
$$CL_i = CL_{pop} \cdot e^{\eta_i} \times \prod_k (1 + CAT_k \cdot \beta_k) \times \prod_l \left(\frac{CON_l}{med(CON_l)}\right)^{\beta_l}$$

- Synthetic population contained 1000 patients with 11 continuous and 5 categorical covariates with various range of correlation; *extreme being BWT-BSA with 0.95 correlation coefficient.*
- Simulated scenarios always included one *low (0.1*) and one *medium (0.3)* covariate effect size on CL.

Simulated covariates (CAT)	Covariate imbalance	Identified covariates	Conclusions
SMK - effect size 0.1 COPD - effect size 0.3	COPD > 5%, SMK > 20%	COPD, SMK	Low effect and high covariate imbalance makes the covariate hard to identify.
	SMK <20%	COPD only	

Simulated covariates (CON)	Sampling	CV limit	Identified covariates	Conclusion
AGE with effect size 0.1 BWT with effect size 0.3	8 samples per patient	<=60%	AGE, BWT	Irrelevant but highly correlated variables (BSA) correctly labeled as irrelevant! High CV impact algorithms ability to identify low effect covariate.
		80	BWT	
	6 samples per patient	<=40%	AGE, BWT (BSA sometimes)	Sparser sampling reduces models ability to identify low effect covariates, as well as increases the chance of selecting wrong (correlated) covariates.
		>=60%	BWT (BSA sometimes)	

Example of training evolution for rich sampling with low CV

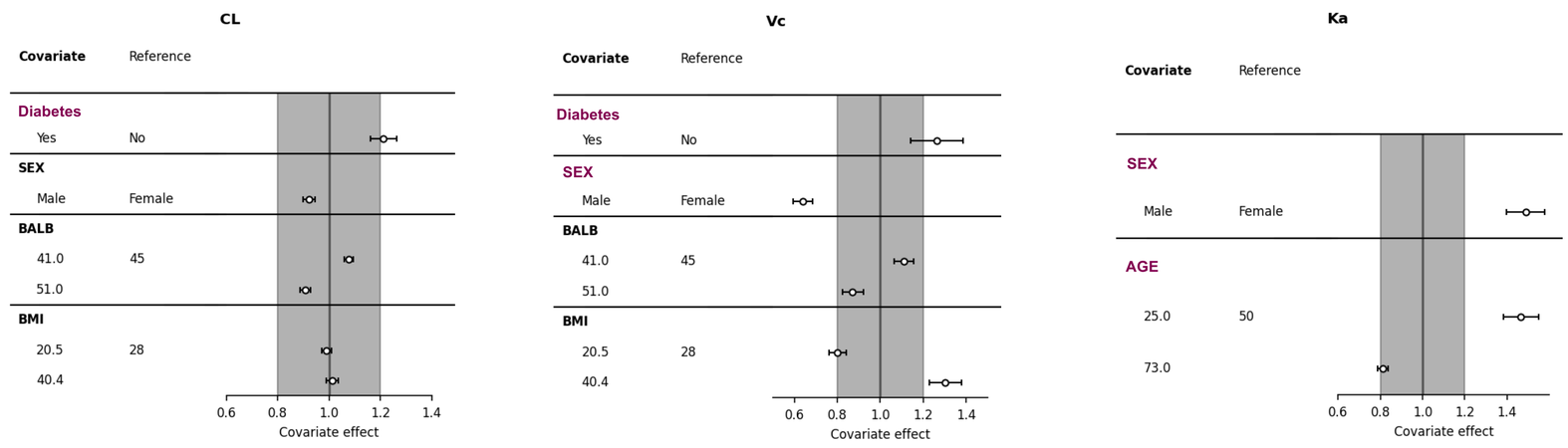


Results: applying the method to AZ real data

- The dataset includes 2527 subjects with a total of 7374 timepoints as described in the PK report [6].
- Structural model is a 2-compartmental model with first-order absorption and first-order elimination.
- Body weight is included in the base model with fixed allometric scaling on CL, Vc, Vp and Q.
- Covariates included in initial screening are:
  - *Categorical*: Sex, Race, Ethnicity, Smoking status, Vaccine status, Diabetes, HIV status, Immunocompromised Flag, Cancer status, Clonal treatment flag, Chronic kidney disease, Chronic obstructive pulmonary disease, Cardiovascular disease
  - *Continuous*: Age, BMI, Height, BSA, ALB, ALT, AST, EGFR, Serum Creatinine, Bilirubin, Creatine Clearance

Comparison of covariates identified by our proposed approach vs. Expert built model				
Model	CL	Vc	Vp	Ka
Expert model	Diabetes	SEX, Diabetes	-	SEX, AGE
Proposed model	SEX, Diabetes, BMI, BALB	SEX, Diabetes, BMI, BALB	-	SEX, AGE

After using NONMEM to fit the proposed covariate model we can use forest plots to further prune the irrelevant covariates based on their clinical relevance (the ones lying within the shaded region), leading to a model very similar to the expert one (purple highlight).



Conclusions

- We have successfully applied neural networks with embedded feature selection layer to identify relevant covariates in both synthetic and real data.
- We noticed that the method performed well under moderately to highly correlated covariates, but low effect size coupled with high CV or covariate imbalance was detrimental for model performance.
- We were able to select subset of relevant covariates on real data, that could further be pruned fitting a full covariate model and lead to a model similar to the expert one.
- These results indicate that the proposed approach can be used to significantly sped up covariate search in PK modelling.

References
1. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN) – a Perl module for NONMEM related programming. Comput Methods Programs Biomed. 2004; 75: 85-94.
2. Fast screening of covariates in population models empowered by machine learning. , Sibieude E, Khandelwal A, Hesthaven JS, Girard P, Terranova N. J Pharmacokinet Pharmacodyn. 2021;48(4):597-609. doi:10.1007/s10928-021-09757-w,
3. Ogami, C., Tsuji, Y., Seki, H., Kawano, H., To, H., Matsumoto, Y. and Hosono, H. (2021), An artificial neural network−pharmacokinetic model and its interpretation using Shapley additive explanations. CPT Pharmacometrics Syst Pharmacol, 10: 760-768. https://doi.org/10.1002/psp4.12643
4. Feature Selection using Stochastic Gates, Yamada, Y., Lindenbaum, O., Negahban, S., & Kluger, Y, Proceedings of Machine Learning and Systems 2020, https://github.com/runopti/stg
5. Lavielle M, Ribba B. Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions. Pharm Res. 2016 Dec;33(12):2979-2988. doi: 10.1007/s11095-016-2020-3. Epub 2016 Sep 7. PMID: 27604892.
6. European Medicines Agency Public Assessment Report for Evusheld, 2022 https://www.ema.europa.eu/en/documents/assessment-report/evusheld-epar-public-assessment-report_en.pdf.

Funding statement: This study was supported by AstraZeneca.
Financial disclosure statement for all authors: All authors were employees of AstraZeneca at the time of study and may own stocks or stock options.