# Losing the Forest: Causal Shapley Additive Explanations for Interpretation of Population-Pharmacokinetic Models

Matthew Wiens, M.A., Elias Clark, Ph.D., Curtis Johnston, Pharm.D.

METRUM RESEARCH GROUP

20 YEARS Anniversary


## Abstract

Forest plots, typically formulated as *Certeris Paribus*, are a commonly used tool to understand covariate relationships in population pharmacokinetic (Pop-PK) modeling. However, these types of forest plots can be misinterpreted leading to unsupported conclusions, particularly, for dosing decisions in clinically-relevant subpopulations. This is often due to the "Table 2 Fallacy" where parameter estimates are conflated with causal effects [1, 2]. To address these limitations, we applied multiple types of Shapley Additive Explanations (SHAP), a tool from the field of interpretable machine learning, to Pop-PK models for comparing conclusions from different methods and visually interpreting model inferences.


## Example 1: Saturable PK

### Methods

- Simulated data from a two-compartment Pop-PK model where clearance is saturable with the half-maximum concentration dependent on EGFR:

$$\mathrm{CL} = \mathrm{TVCL} \times \left(\frac{\mathrm{WT}_i}{\mathrm{ref}_{\mathrm{WT}}}\right)^{0.75} + \frac{\mathrm{CL}_{\mathrm{MAX}}}{\mathrm{CL}_{50} + \mathrm{CENT}}$$

$$\mathrm{CL}_{\mathrm{MAX}_i} = \mathrm{TVCL}_{\mathrm{MAX}} \times \left(\frac{\mathrm{WT}_i}{\mathrm{ref}_{\mathrm{WT}}}\right)^{0.75}$$

$$\mathrm{CL50}_i = \mathrm{TVCL}_{50} \times \left(\frac{\mathrm{eEGFR}_i}{\mathrm{ref}_{\mathrm{eEGFR}}}\right)^{\mathrm{COV}_{\mathrm{eEGFR,CL}_{50}}}$$

- Allometric scaling on other flows and volumes
- Simulated dosing every 12 hours, with negligible accumulation
- Analyzed area under the concentration-time curve (AUC) as a summary exposure metric

### Results

- SHAP analysis identified differing effects of eGFR depending on weight and much larger effects when eGFR and weight are on the tails of the dataset

**Figure 1: Forest Plot.** Traditional forest plot shows a small effect of eGFR (3%), and a total weight plus eGFR effect of up to 20% (16% + 4%), relative to a reference patient with a weight of 70 kg and 86 mL/min/1.73m$^2$

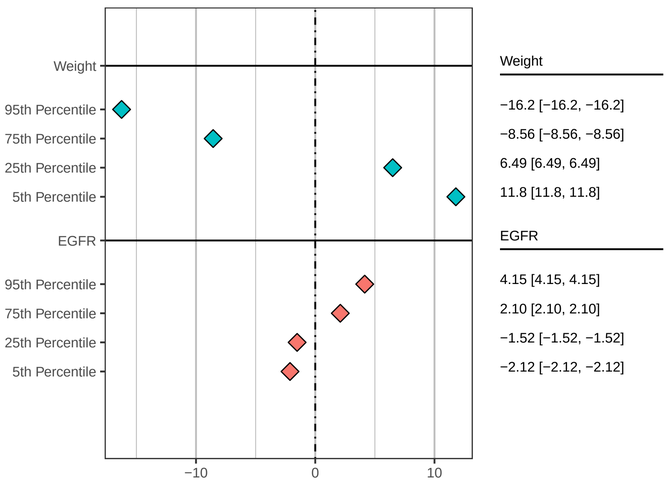
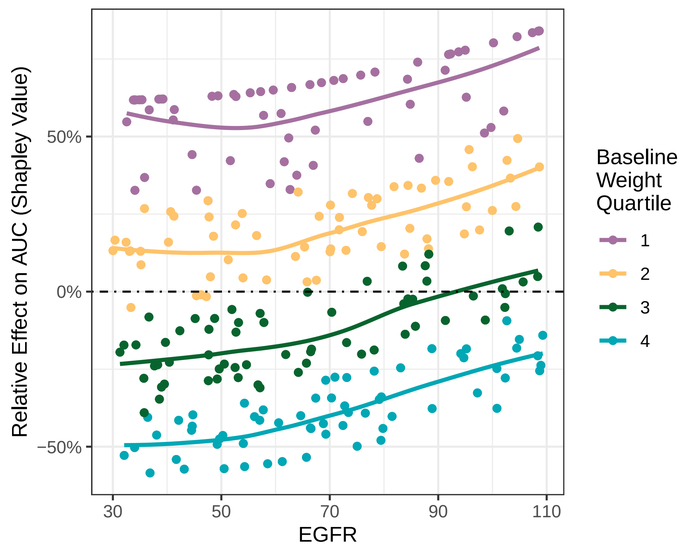



**Figure 2: SHAP Covariate Analysis** There was an interaction between weight and EGFR, showing much lower AUCs for high weight. Relative differences were up to 65%,compared to the reference. In addition, there is non-linearity in the effect of EGFR.




## Example 2: Causal Covariates

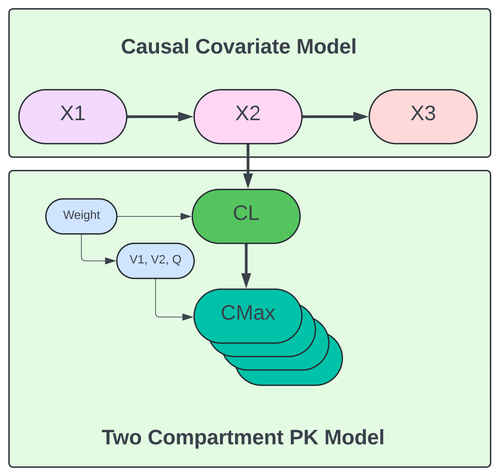
### Methods

**Figure 3: Covariate and PK model. Only one covariate, $X_2$, enters the PK model.**

- Three covariates were simulated, each standard normal but with a specific causal dependence structure:

$$\mathbb{E}[X_1] = 0$$
$$\mathbb{E}[X_2] = 0.3 \times X_1$$
$$\mathbb{E}[X_3] = 0.8 \times X_2$$
$$\log(\mathrm{CL}) = \log(\mathrm{TVCL}) + 0.75 \times \log\left(\frac{\mathrm{WT}_i}{\mathrm{ref}_{\mathrm{WT}}}\right) + \frac{X_2}{2} + \eta_{CL}$$

- Data was simulated from a 2-compartment PK model, with 1 input covariate ($X_2$) and allometric scaling for weight
- **Causal Shapley Values** [2, 3] and **population simulations** were used to analyze differences across the population in the summary exposure metric of Cmax




### Results

- The analysis methods provide different conclusions for which covariates impact Cmax
- Several plots and associated smooths (blue lines) are "correct," each to support different conclusions
- Asymmetric causal Shapley Values best reconstructed the known causal structure
  - *Direct effect*: Holding all other variables constant, the difference in Cmax for varying this covariate
  - *Total effect*: The effect on Cmax from varying this covariate

**Figure 4: Marginal Shapley Values.** Shows predictive effects only and splits the effect between the potential covariates.

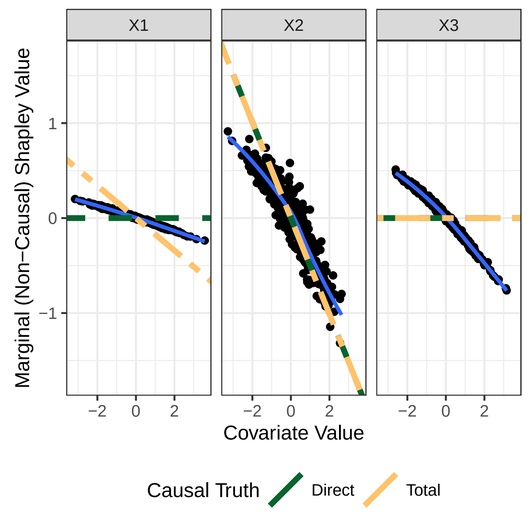



**Figure 5: Asymmetric Causal Shapley Values.** Correctly identifies total causal effects.

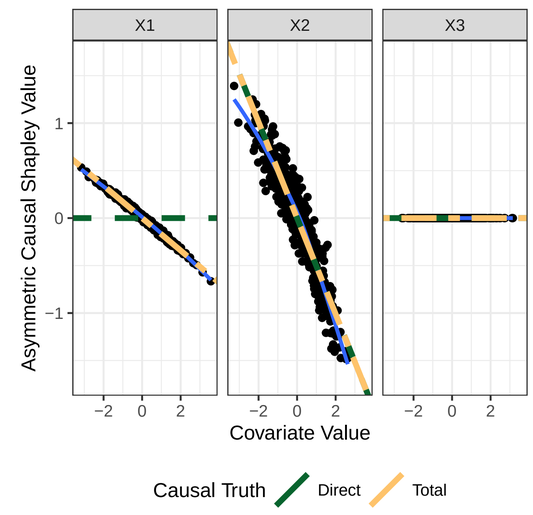



**Figure 6: Parameter Values Forest Plot.** Parameter estimates in the simulated model, visualized as a forest plot.

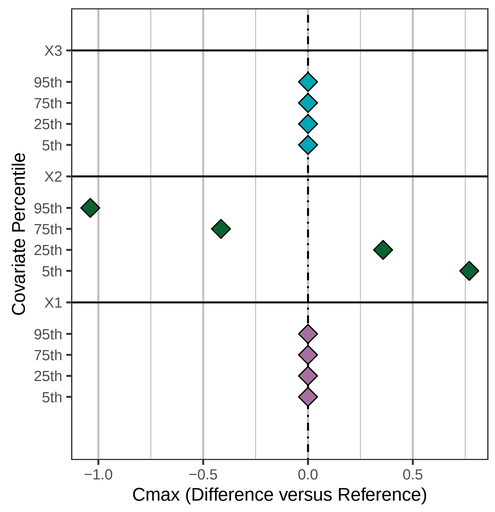
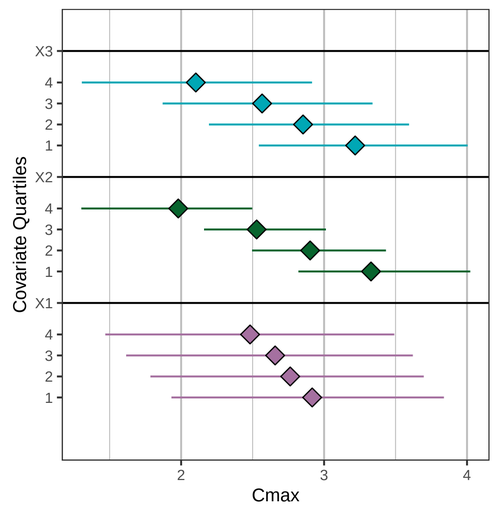



**Figure 7: Population Simulation.** Shows predictive effects and 90% prediction intervals within subpopulations. Covariates that do not have a causal effect can still have different distributions of Cmax across quartiles and, therefore, can be used for population adjustment.




## Metrum Research Group Publications and Posters

## References


[1] Westreich, D. and Greenland, S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am. J. Epidemiol.* **177** (2013):292–298.

[2] Heskes, T., Sijben, E., Bucur, I.G. and Claassen, T. Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models. *arXiv [cs.AI]* (2020).

[3] Lundberg, S.M. and Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)* (2017).


## Conclusion

- Shapley values can be used to analyze PK and PKPD models in addition to black-box AI/ML models, and support conclusions beyond those supported by forest plots:
  - Identification of potential meaningful subpopulations and interactions which univariate forest plots may not identify
  - Support identification of correlation versus causation within complex models and causal assumptions
  - As a technique to bridge between AI/ML modeling strategies and established PKPD NLME approaches
- In the presence of causal dependence in covariates, parameter estimates (and forest plots of parameter estimates) were not sufficient for making decisions about dose adjustments
  - Features used for dose adjustments do not need to be included in a model nor have significant effects
- The analysis question of interest is critical for choosing appropriate model visualizations to assess

Presented at the American Conference on Pharmacometrics; 10 November - 13 November 2024

Copies available at: www.metrumrg.com/all-publications