



SHAP analysis provides a
**universal,
flexible, and
scalable**
alternative to traditional
sensitivity analysis



CONCLUSIONS

Machine learning methods are underutilised within health economics

Neural network-based metamodels replicate complex health economic models with speed and accuracy

SHAP analysis reveals hidden drivers and interactions within neural network outputs

Combining neural network-based metamodeling with SHAP analysis provides modellers with a toolkit for scaling complex models and untangling the key drivers of model outcomes

INTRODUCTION

- Neural network-based metamodels can effectively replicate the outcomes of complex health economic models¹
- SHapley Additive exPlanations (SHAP) analysis is a game-theoretic approach that can be used to interpret machine learning model output, including neural networks
- SHAP analysis highlights both overall feature importance and each feature's contribution to individual predictions, uncovering complex interactions among model parameters

SCALING COMPLEX MODELS

SHAP analysis complements neural network-based metamodeling to unearth model interactions and sensitivities

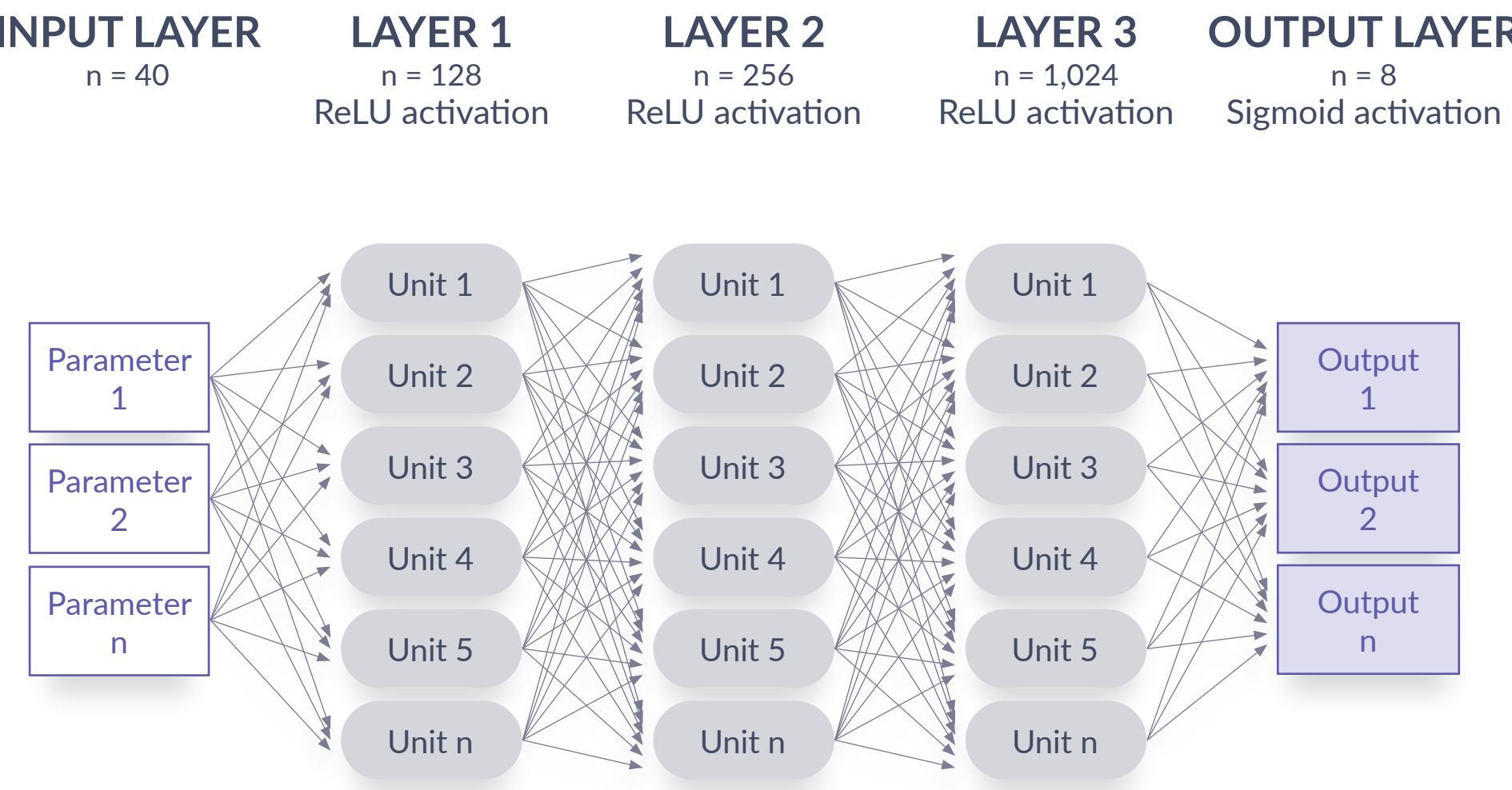
OBJECTIVES

- This study demonstrates the use of SHAP analysis to interpret predictions from a neural network-based metamodel of chronic kidney disease (CKD) prevalence, revealing relationships between parameters, key outcome drivers, and sources of uncertainty

METHODS

- We previously demonstrated that metamodels can be utilised in health economics to bypass complexity and democratise modelling¹
- A neural network-based metamodel (Figure 1) was fitted to a dynamic prevalence model of CKD, predicting a variety of population outcomes between 2024 and 2060, based on a system of ordinary differential equations (ODEs)
- Here, we retrained the model to predict eight output categories for the year 2060 and used Latin hypercube sampling (LHS) to generate training and test sets of 20,000 and 10,000 model scenarios, respectively. The outcomes include total population, total excess deaths, total end-stage kidney disease (ESKD) related deaths and total numbers of conservative care, peritoneal dialysis, haemodialysis, home haemodialysis and transplant cases
- SHAP analysis was then carried out on the metamodel to quantify the influence that varying treatment eligibility proportions for 40 different age and disease-stage combinations has on the eight model outcomes
- We generated a range of heatmaps, bee-swarm, and waterfall plots to visualise SHAP values for different input values, outcomes, and scenarios

Figure 1. Schematic diagram of the neural network-based metamodel trained to replicate the selected results of the CKD dynamic prevalence model



LIMITATIONS

- Computing SHAP values may require considerable computational resources and time, subject to the underlying model structure and the number of input-output scenarios that would need to be explained
- Sampling inputs for training scenarios must be carefully considered, as the parameter space needs to be captured as much as possible. While LHS ensures efficient coverage, high-dimensional input spaces would require an increased number of training scenarios, which in turn impacts the required hardware for metamodel training

RESULTS

SHAP analysis provides a global view of feature importance

- Performing SHAP analysis on the neural network-based metamodel of CKD prevalence reveals the relative feature importance for each outcome of interest. Figure 2 presents a bee-swarm plot illustrating the relative influence of input features (i.e. the number of people of a particular age or disease state eligible for treatment) on eight key model outcomes in the year 2060, including the size of the CKD population, the number of excess deaths, ESKD-related deaths, and patients receiving haemodialysis or transplantation. Each scenario is represented as a coloured dot, with an associated feature value

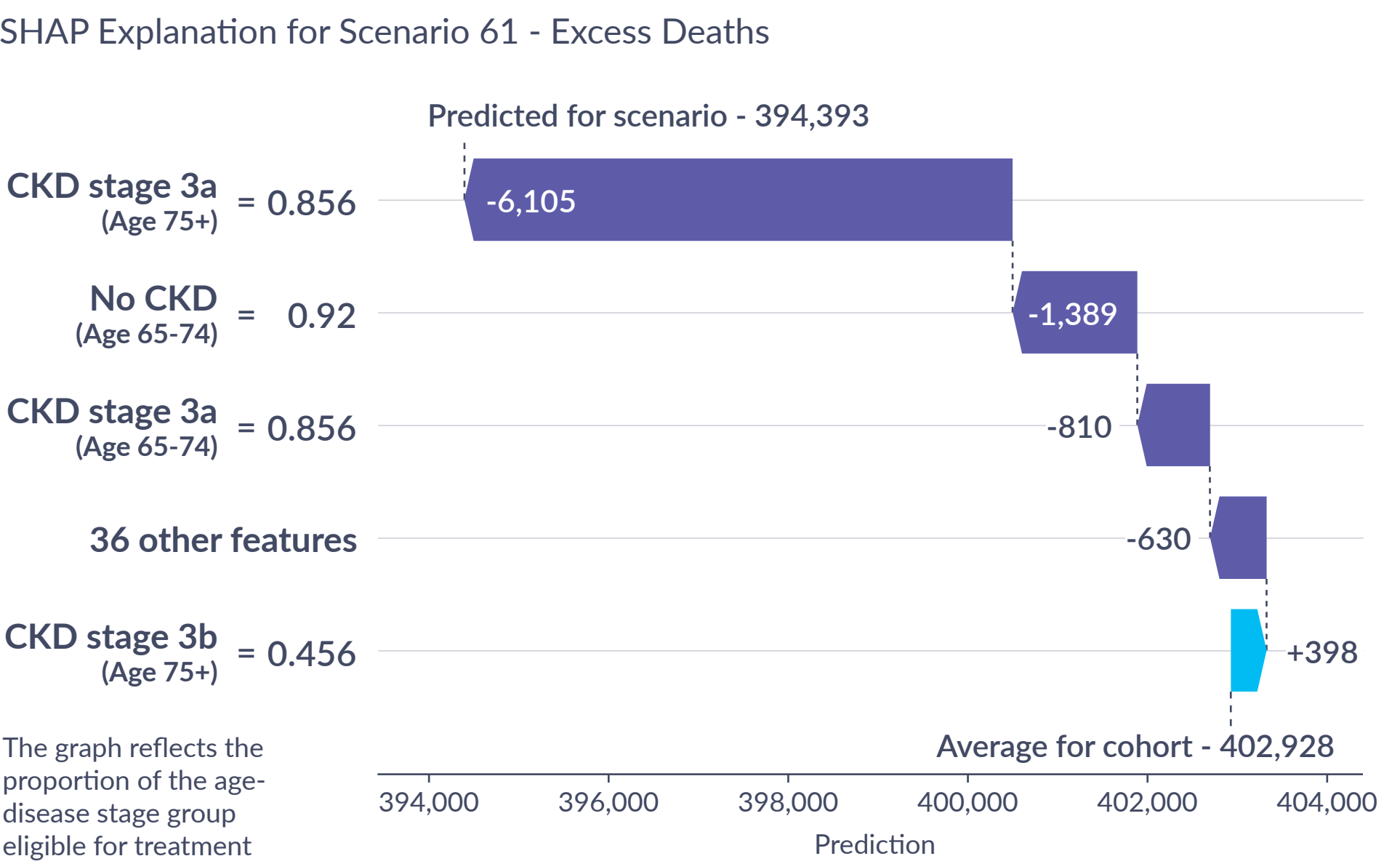
Figure 2. SHAP values for different model features by model outcome. Features are ordered by their importance to the outcome of interest



Interrogating the drivers of model predictions

- SHAP analysis also reveals the drivers of individual model predictions. Figure 3 shows a waterfall plot for a randomly chosen model scenario, where predicted excess deaths are significantly lower than the average across all scenarios. Yellow arrows indicate features that increase the predicted value, while maroon arrows denote those that decrease it
- The mean expected value of excess deaths is shown at the bottom, with an influence of each model parameter affecting the final prediction and the magnitude of effect. Since input parameters relate to the proportion of people eligible for treatment in the dynamic model (ranging between 0 and 1), the mean value of each input parameter is 0.5
- We can see that reducing the proportion of people who are eligible for treatment aged 75+ within the CKD stage 3b has a small increased effect on excess deaths, while increasing the proportion of those people aged 75+ with CKD stage 3a has the largest impact on the reduction in excess deaths in the given scenario
- This highlights the ability of SHAP analysis to provide local interpretability to individual scenarios, something that traditional sensitivity analysis struggles with

Figure 3. Waterfall plot for a randomly selected model scenario for predicted excess deaths



Uncovering interaction effects and distributional insights

- SHAP heatmaps allow for the exploration of feature interactions and distributional patterns across scenarios, capturing heterogeneity in effects
- By arranging the outcome value in increasing fashion and overlaying individual input scenarios in a heatmap in the order of feature importance, modellers can identify non-linear relationships and interactions between parameters. This approach can further be extended to identify interactions between specific inputs of interest and explore how they interact across the permissible input space
- In the CKD model, clustering reveals that combinations of features drive higher predictions of excess deaths. In Figure 4, we can see that very low input values for those aged 75+ in the 'No CKD' and 'CKD stage 3a' disease groups are important drivers of higher excess death predictions

Figure 4. Input values heatmap: features are ordered by model importance and SHAP value for excess deaths. Each line represents a model scenario with an associated input value

