University Hospital rechts der Isar
Technical University of Munich

Published in Frontiers in Pharmacology:
kurzelinks.de/wendl2022frontiers

Identification of target groups and individuals for adherence interventions using tree-based prediction models

Johannes Wendl

Institute of General Practice and Health Services Research, School of Medicine, Technical University of Munich
Johannes.wendl@mri.tum.de

Background

Adherence to medication of chronically ill patients is defined as the extent to which a patient follows agreed recommendations from a health care provider.[1] It can be crucial for treatment success and hence can decrease future health costs. In some populations regression models do not show this relationship overall.[2] Therefore, we aim to estimate subgroup-specific and personalized effects to identify target groups for interventions.

Methods

We used German claims data collected between 2012 and 2015 to define a cohort of diabetes type 2 patients with corresponding drug fills (n=85,162). Outcome was mean annual total costs in 2013-2015. All other variables were measured in 2012. Our main predictor, adherence, was operationalized by proportion of days covered (PDC) by diabetes medication.

$$\text{PDC} = \frac{number\ of\ days\ covered}{number\ of\ days\ under\ medication} \times 100$$

Covariates were age, sex, Charlson's Comorbidity Index (CCI), initial total costs, three-level severity based on prescription guidelines and fills and surrogates for health behavior (participation in disease management programs and influenza vaccination). In models with split variables we used age, CCI, initial total costs and severity. We estimated three kinds of models using training data (50%):

1. **Linear Regression Model:** Estimation of overall adherence effect
2. **Model-Based Trees:**[3] Identification of subgroups and estimation of subgroup specific adherence effect
3. **Model-Based Random Forests:**[4] Estimation of patient individual adherence effects by similarity weighted regression models

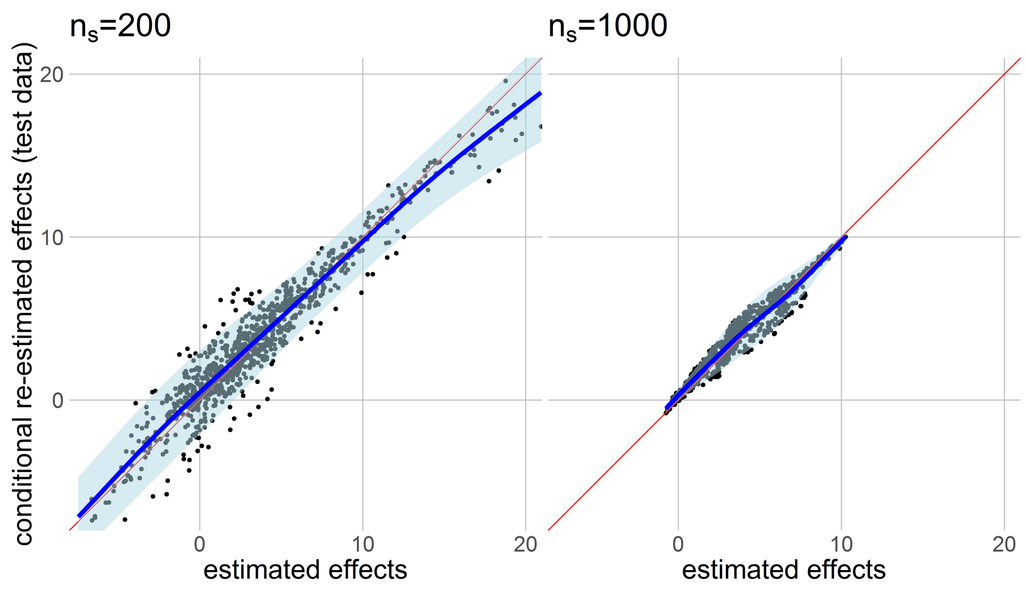
To assess the performance of the latter, we conditionally re-estimated the personalized adherence effects using test data, the fixed structure of the forest, and fixed effect estimates of the remaining covariates of the model. For each patient we compared the effect estimated by the forest to the conditional one re-estimated on test data in a calibration plot. A GAMLSSS regression of these two estimates with 95% prediction interval was used to identify patients where we can expect a negative adherence effect with the given certainty based on the forest's respective estimate.
Different specification of minimal terminal node size (200 and 1000) defined the extent of the desired similarity between observations.

Results

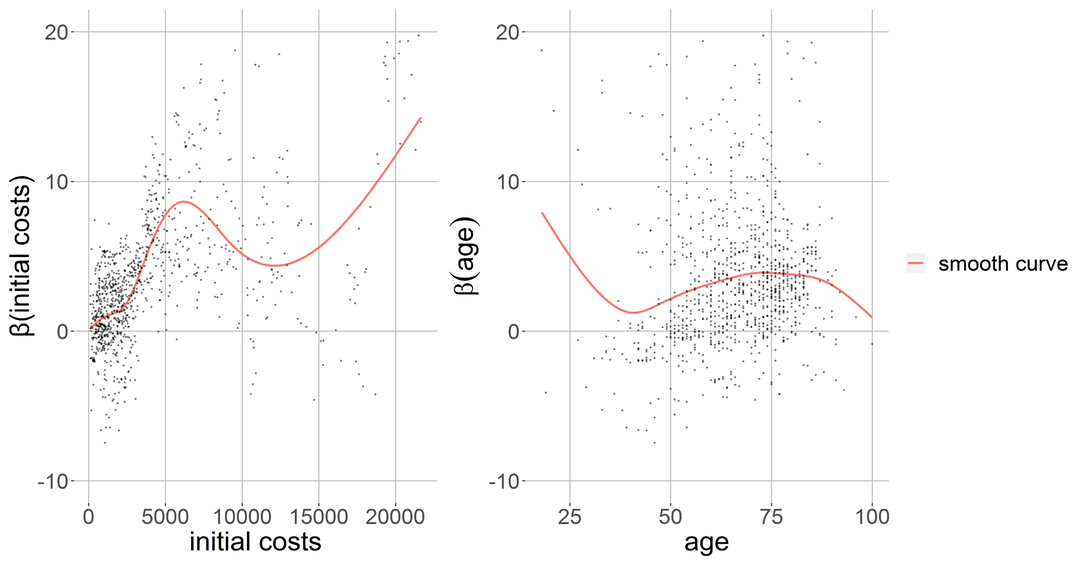
Our simple linear regression model estimated a positive adherence effect on total costs of 3.92€ per PDC-point (p≤0.001). In contrast our model-based tree detected five subgroups including one of 24% of patients with a negative estimated adherence effect of -1.69€ per PDC-point.

Subgroup	Estimate	P-value	n (%)
initial costs <= 3130 & age <= 63	-1.69	0.087	10,394 (24.4)
initial costs <= 3130 & age > 63	1.86	0.069	14,154 (33.2)
initial costs <= 7307 & initial costs > 3130 & age <= 76	9.17	0.000	7,930 (18.6)
initial costs <= 7307 & initial costs > 3130 & age > 76	1.60	0.612	2,851 (06.7)
initial costs > 7307	6.21	0.012	7,252 (17.0)

Our model-based random forest estimated negative individual adherence effects in 21% of patients ranging from -7.45 to -0.01€. Even when adding the 95% prediction interval (light blue area) we can expect a negative adherence effect with the given certainty in 4% of patients when the estimated effect of the forest was lower than -3.21€.



The calibration plot, where perfect precision is illustrated by a diagonal red line, shows that the effect estimates where well calibrated. The forest with smaller terminal node size had a higher between-person variance of effect estimates but lower precision.



Partial dependence plots show the relation of the personalized effect estimates to the split variables. They increase continuously by initial costs until around 6000€ and also in the main age groups between 50 an 80.

Discussion

While a simple linear regression model showed positive association between adherence and costs, model-based trees as well as forests identified patients with negative effects. Different variance and precision of estimates depending on forest specification of terminal node size indicate a respective trade-off.
It shows that tree-based models can identify patients with different effects up to the individual level. With our calibration-like approach, we were able to asses the quality of the effect estimation by model-based random forests.
The identified patients can be assigned to target groups for adherence-promotion interventions with the aim to increase health and decrease associated costs. The proposed method can also be applied to predict other outcomes such as hospitalization risk to maximize positive health effects of an intervention.

References

1. Sabaté, E., & Sabaté, E. (Eds.). (2003). Adherence to long-term therapies: evidence for action. World Health Organization.
2. Cutler, R. L., Fernandez-Llimos, F., Frommer, M., Benrimoj, C., & Garcia-Cardenas, V. (2018). Economic impact of medication non-adherence by disease groups: a systematic review. BMJ open, 8(1), e016982.
3. Seibold, H., Zeileis A., and Hothorn T. (2016). Model-Based Recursive Partitioning for Subgroup Analyses. The International Journal of Biostatistics 12 (1): 45–63.
4. Seibold, H., Zeileis A., and Hothorn T. (2018). Individual Treatment Effect Prediction for Amyotrophic Lateral Sclerosis Patients. Statistical Methods in Medical Research 27(10): 3104–25.