



Best Practices: Analyzing classical clinical trial data with ML methods

Uwe Pritzsche
Data Scientist
HMS Analytical Software GmbH

Agenda

- Motivation
- “SAP vs. AI”
- Pooled vs. Single-Study
- ADaM specialties
- Methodology: EDA (explorative data analysis), Model comparison and Winner Model Evaluation, Interpretation by XAI (explainable AI)

Motivation

- Clinical:
 - Increase insights from clinical trial utilizing AI/ML
 - AI/ML-based identification of predictive marker for efficacy, safety and tolerability endpoints for treated patients
 - → Possibly improve therapy of patients
- Methodological:
 - Establish new methods/algorithms for applying AI/ML on clinical study data

“SAP vs. AI”:

Classical statistical analysis vs. ML based predictive modeling

SAP	AI
Report primary and secondary endpoints (e.g. PFS)	Find prognostic factors for many endpoints (e.g., PFS, Discontinuation, Time-to-sAE, ...)
Additionally, report lot of GCP-related information (e.g., AE listings)	Concentrate on analytically relevant information (e.g., prior_therapy highly predictive for xAE)
Check prespecified hypothesis (of association) by statistical significance	Check predictive capability of baseline information → Discover unknown pattern
Analysis is highly limited due to regulatory and historical reasons (GCP)	Exploratory approach using modern analytical methods
Goal: Get approval from FDA	Goal: Derive new insights which can be utilized in future trials
At most “Longitudinal Mixed Model” with a few, prespecified predictors applied for one endpoint, usually only checking for correlation but not prognosis	Machine Learning Model with all information available at baseline in order to predict various endpoints

Pooled vs Single-Study:



	Pooled	Single-Study
Missing data	Some data does not exist for all studies due to different CRFs → Need to create some separate models anyway	Do not need to distinguish which model can run on all or just a subset of the data & SAP can directly serve as “domain expert” without any additional checks of other SAPs
Complicated data	Some information was not integrated due to lack of awareness of importance for ML modeling and might be complicated to integrate → Need ML data model first, then integrate	All information which exists for single study is directly available
Interpretation Obstacles	High potential for interpretation obstacles due to confounding by study	No “study” confounding (and validate insights on other studies on a qualitative level)
Signal Strength	Some predictors do not make sense for all 4 studies -> Full model would blur signal	Predictors are derived study-specific and therefore cannot blur any signal
Data dimension advantages	More records = longer Analysis-table → Predictive signals are sharper	More columns = wider Analysis-table → More predictive signals which can be qualitatively validated on other studies
Analysis data format	Data is just in SDTM format → Potential of different derivation of endpoints which ends up in different results to original SAP based analysis → higher effort of derivation of important endpoints	Usually there are already ADaM datasets which is the basis of the Analysis-table (and will be enhanced by additional derivation of lacking information) from basis data → no difference to SAP based analysis & utilization of derivation of important endpoints
Standard data format	Already in SDTM format	Might not be in SDTM format and ADaM conform analysis datasets might not exist
Data prep for analysis	Just one-time	For every study
Delays	Need to wait for final integrated data	Can be done immediately on every study

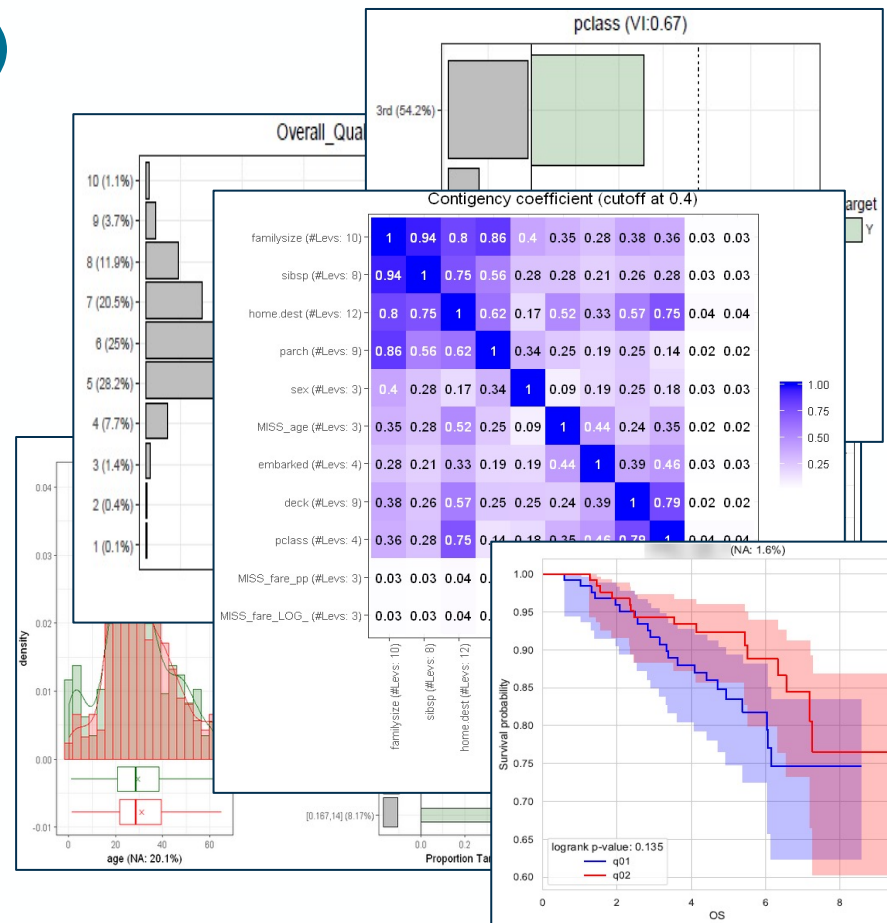
ADaM specialties

- Lots of possible targets to predict:
Response rates (orr, cbr, pfs, death), **patient reported outcomes** (xxx_deterioration/improvement with xxx=QoL, social_functioning, ...), **safety** (ae_serious/severe, ae_xxx_severe with xxx=important AE), **tolerability** (study_discontinuation, treatment_discontinuation, ...)
- Lots of possible predictors/features:
Demographics&Vitals (cohort, age, ecg, sysbp, diabp, ...), **MH** (mh_xxx with xxx=important MH), **disease-specific** (e.g. histological_grade, days_since_primary_diagnosis, ...), **prior therapy specific** (prior_xxx with xxx=important prior therapy, ...)
- Important: All predictors measured at baseline in order to prevent data leakage (“information from the future”)! → “prognosis, not correlation”
- Time2Event/Survival modeling: Model time as discrete with 1 record per time-point and patient
→ Can treat time2event targets as classification problem (i.e. logistic regression) and apply any ML based classification algorithm (Likelihoods coincide with Cox Non-Proportional Hazard model)

Methodology:

1. EDA (explorative data analysis)

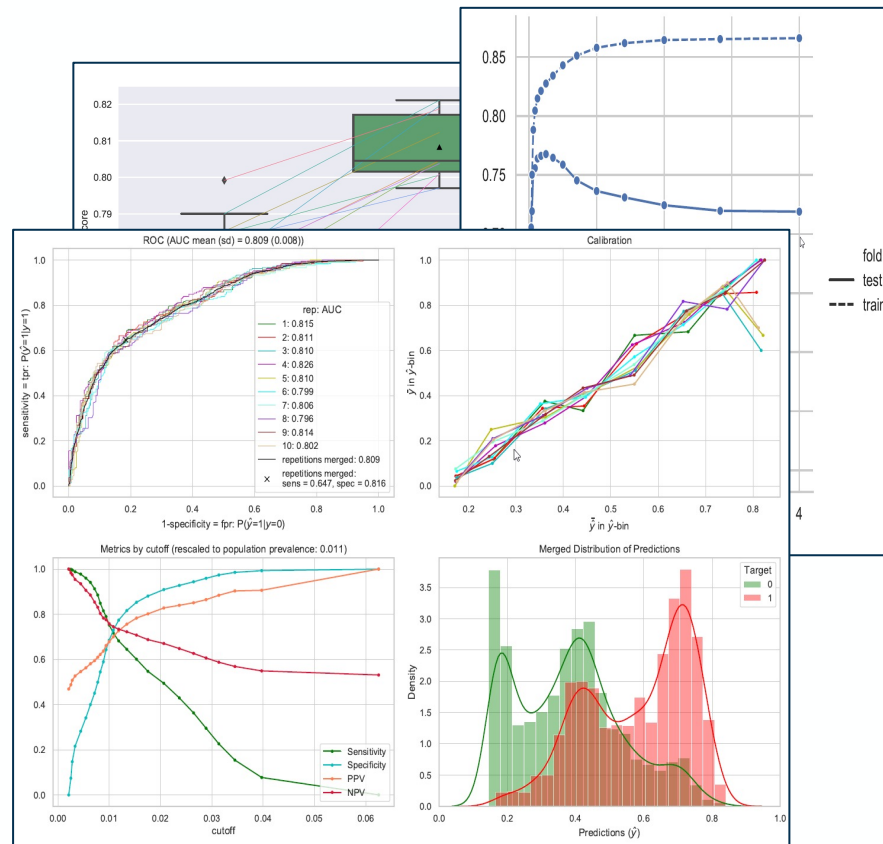
- Plot, plot, plot = No tables and listings, only figures
- Investigate each feature regarding outlier, unusual distribution, missings, high correlation with other features, unusual correlation with target (must be done for each target investigated)
- Depending on the result of this investigation appropriate preprocessing steps must be conducted, like outlier removal e.g. by winsorizing, feature dropping due to high missing frequency, imputation, log-transformation for right-skewed features, bucketing high cardinality categorical features, ...
- Important: Do not use for feature selection → “data snooping”



Methodology:

2. Model comparison and winner model evaluation

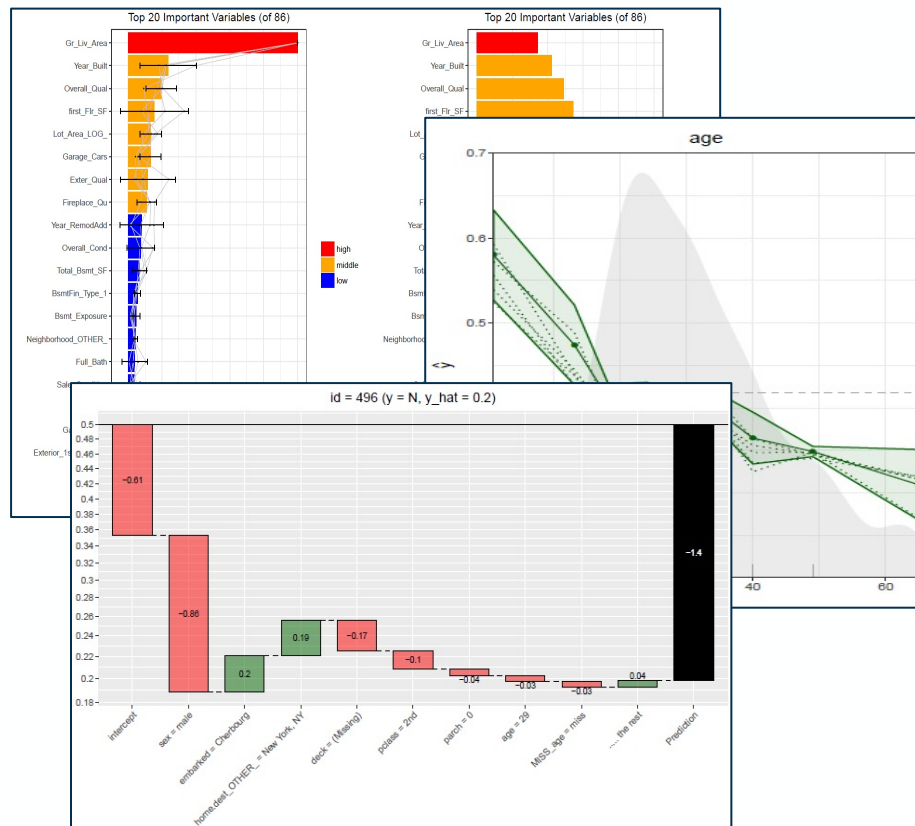
- Algorithms (to compare), at least:
 - Elastic Net / LASSO = Penalized Regression: Fully interpretable, robust against correlated or irrelevant features, low overfitting potential
 - Boosted Trees: Mostly interpretable (by XAI), robust against correlated or irrelevant features, implicit modeling of interaction effects, higher overfitting potential, best algorithm for tabular data (regarding predictive performance)
 - Cross validate results and take care of overfitting (“mind the generalization gap”)
- Plot and discuss predictive performance, regarding discrimination and calibration (“risk class accuracy”), of winning algorithm for each target-model
- Important: Evaluate successful models on different trial



Methodology:

3. Interpretation by XAI

- Variable importance:
Importance of each predictor regarding predictive performance, possibly crossvalidated (instead of p-values and significance)
- Partial Dependence Plots:
Effect of each predictor on target on average (multivariate adjusted), equivalent to the “betas” of regression
- Shapley Values: Individual effect contribution of features on target for a single patient





Thank you for your attention

uwe.pritzsche@analytical-software.de