

Handling dropouts in longitudinal observational studies – an illustration of a workflow from data preparation to model building and model diagnostics in Alzheimer’s disease

Giulia Lestini¹, Etienne Pigeolet¹, Neva Coello¹, Martin Fink¹, Thomas Dumortier¹

¹Novartis Pharma AG, Basel, Switzerland

Background

Longitudinal observational studies collecting cognitive data can be used to model Alzheimer’s disease progression. Those models cannot be adequately evaluated using standard simulation-based diagnostics given the high dropouts rates as commonly observed in observational studies, even if dropouts happen “at random” [1]. The solution is to reproduce the dropout process in the simulation-based diagnostics of the longitudinal cognitive score. For that a dropout model is needed

Objective

To describe a workflow for developing a model of the dropout process using a time to event approach

Challenge

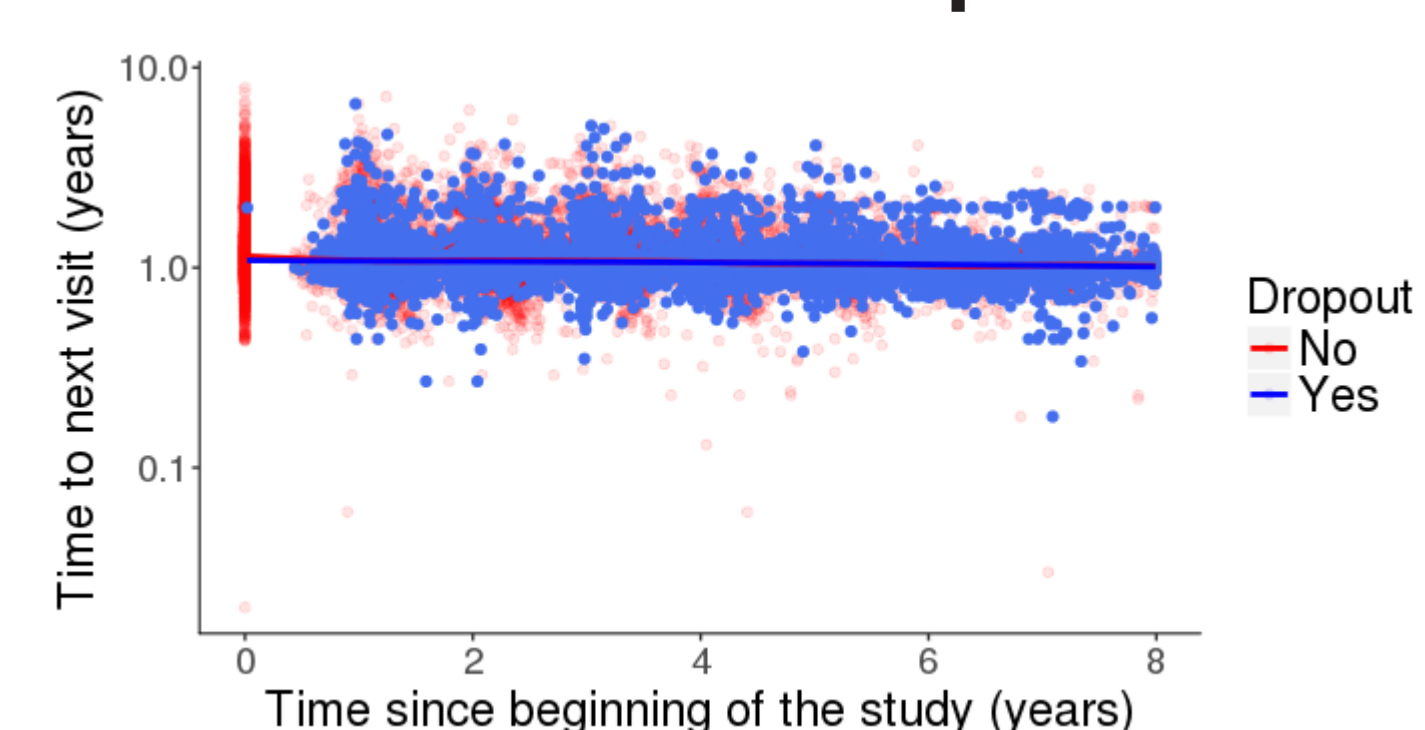
- Dropout time is unknown: it can be at any time between two subsequent visits
- Subsequent visit time is irregular overall and unknown for dropout individuals

Data

Our work is based on longitudinal data collected from three cohort studies (ROS, MAP and MARS) of memory and aging at the Rush Alzheimer’s Disease Center.

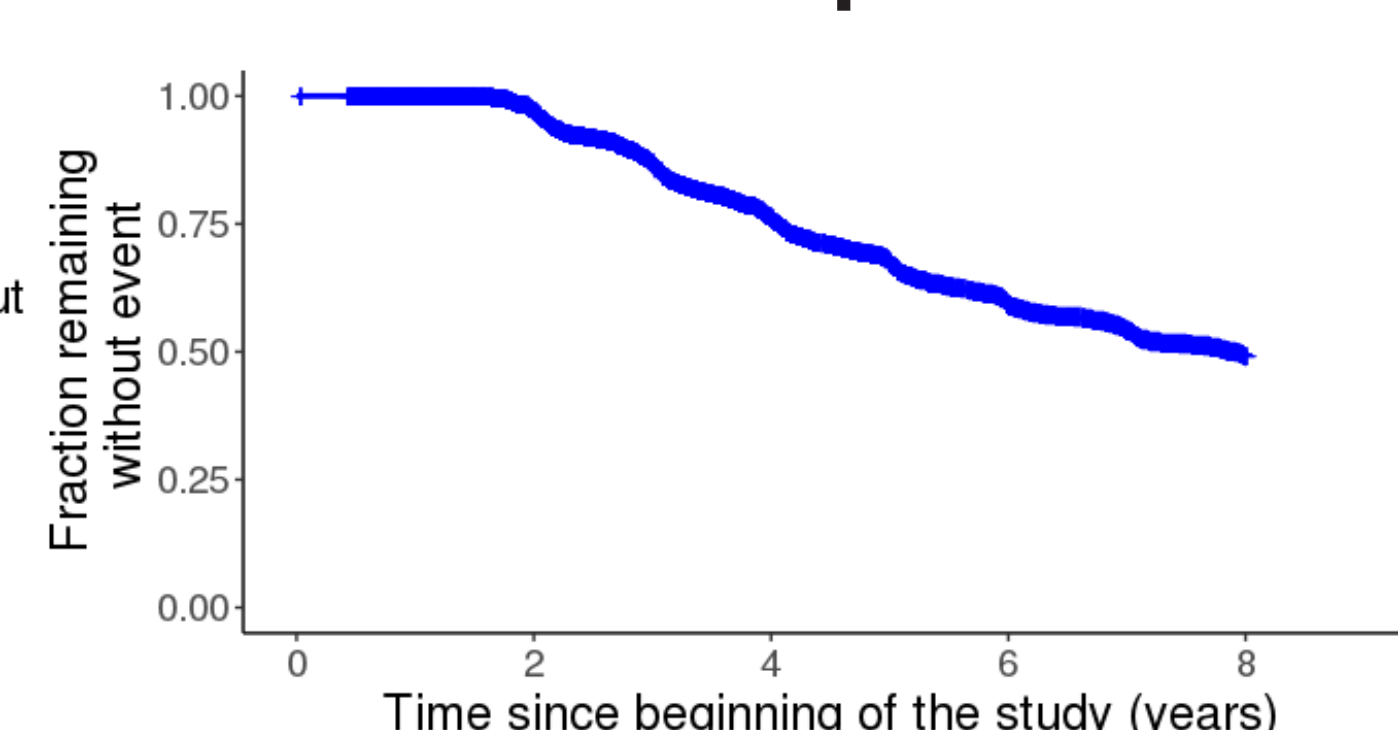
- We apply a cut-off at 8 years and we define as dropouts those individuals with their last cognitive assessment occurring before year 8. Individuals who have assessments up to 8 years or beyond are right-censored at 8 years.
- We exclude subjects with baseline cognitive score not available «n.a.» and records with longitudinal cognitive score «n.a.»
- In this analysis data set 1124 of 2194 subjects drop out before year 8.

Figure 1: Time-to next visit in individuals who dropout* or not



*For dropout individuals, time to next visit is imputed (see method section)

Figure 2: Kaplan Meier* of time to dropout



*Kaplan Meier adapted to interval censoring, with imputed end of censoring interval using the stochastic algorithm (see method section)

Method

- We propose to model the dropout times using time-to-event analysis with interval censoring as the dropout can occur at any time between two subsequent visits
- Since visits are not scheduled at the same time for all subjects, we stochastically impute the end of the censoring interval based on the information available from subjects having a subsequent visit (Figure 1)
- Two-step approach of model-building:
 - Semi-parametric Cox model to investigate the shape of the baseline hazard
 - Parametric time to event models to allow more flexibility in the functional relationship between the covariates and the hazard. Baseline covariates included: age and cognitive score

Use of diagnostics such as time-independent Martingales residuals and Kaplan Meier of the Martingale (predicted cumulative hazard) [2] are produced to assess the quality of the models

Conclusions

The developed dropout model – based on our presented workflow – can be used to replicate the dropout process in the simulation-based diagnostics of the longitudinal cognitive score. The algorithm of stochastic interval imputation provides a good approximation of the end of the censoring interval for dropout subjects. The appropriateness of the functional relationship between baseline covariates, age and cognitive score, and dropout hazard is shown using Martingale residuals.

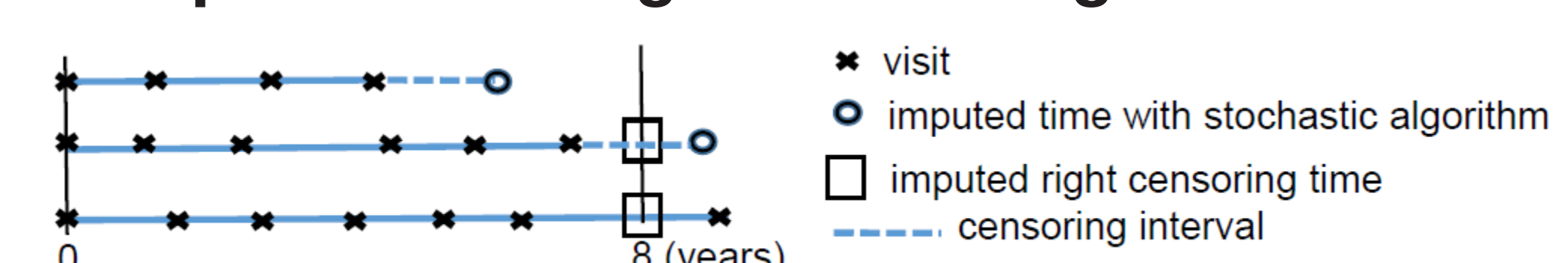
Results

Stochastic imputation algorithm

The algorithm to stochastically impute the end of the censoring interval for a subject who drops out consists of the following steps:

1. Calculate (for every id) the time to last observed cognitive score «TTLOS»
2. Define as dropout the individuals with TTLOS<8
3. For the dropout-subjects stochastically impute the «next expected visit»
 - From subjects having a subsequent visit identify the time to the next known follow-up visit («follow-up interval»)
 - For dropout individuals impute a putative follow-up visit by randomly sampling from the «follow-up intervals» (Figure 1) – only considering visits happening +/-0.5 years around the dropout-individual’s last visit.
4. For all visits (imputed or not), the first visit after 8 years is right censored at 8 years, and subsequent records (with years >8) are removed

Figure 3: Imputation of right censoring time in 3 individuals



Model building

Cox proportional hazard model

The cumulative baseline hazard obtained from the Cox proportional hazard model (Figure 4) suggests that the hazard of dropping out in the first 1.5 years is almost zero, and then it constantly increases over time.

Figure 4: Cumulative baseline hazard - Cox model

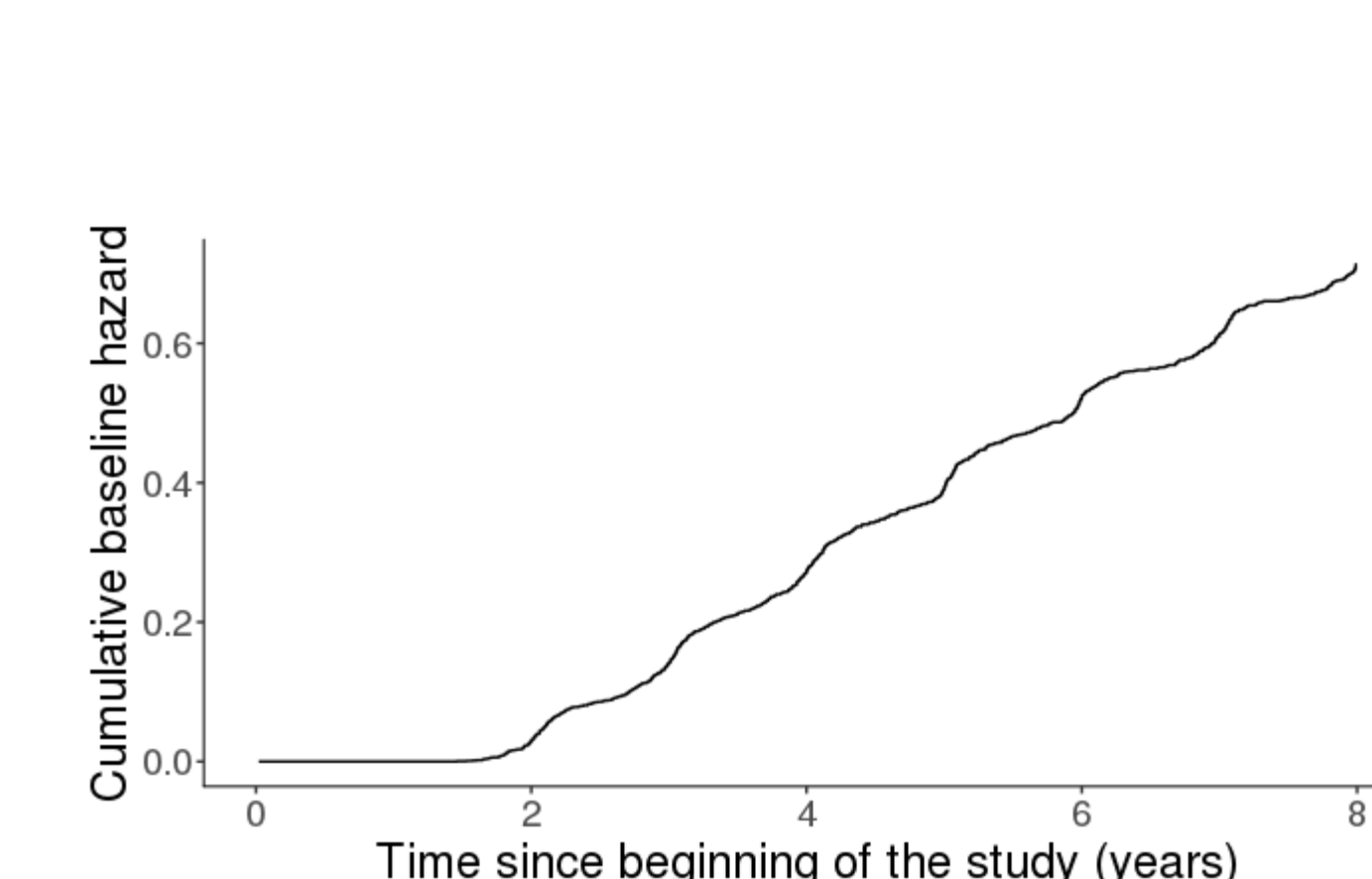
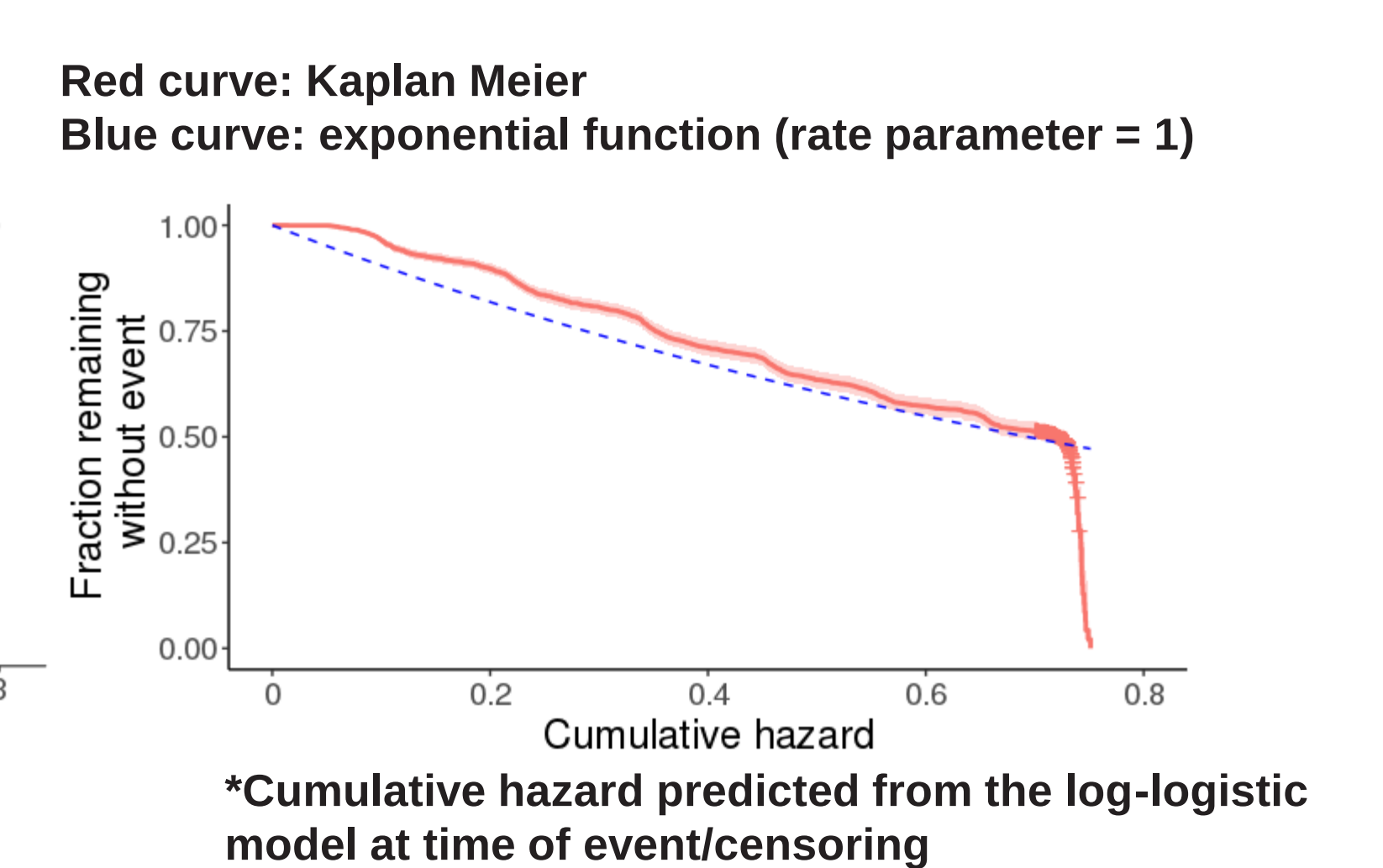


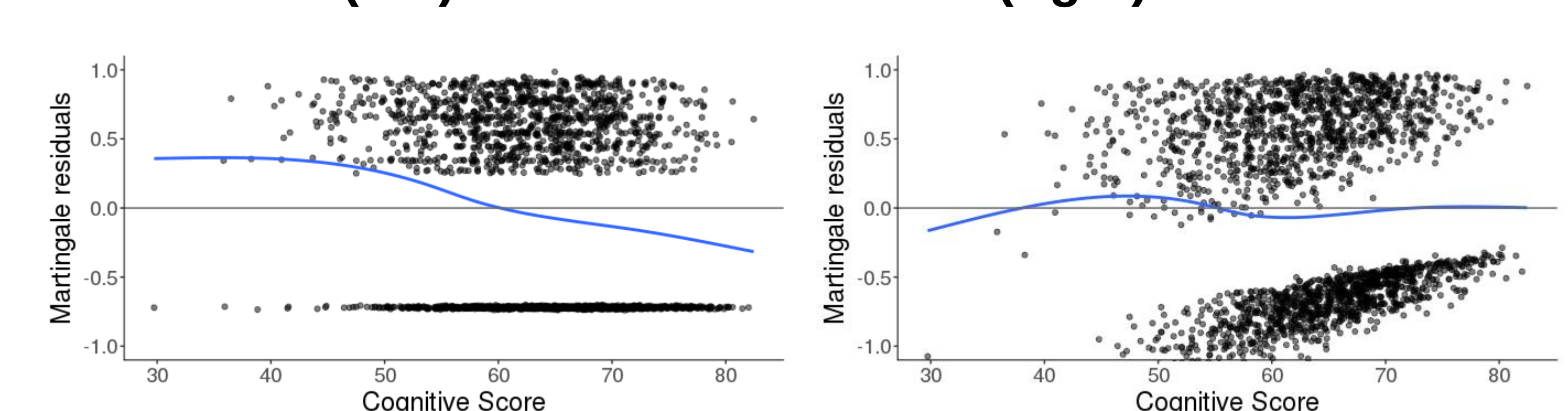
Figure 5: Kaplan Meier of cumulative hazard with covariate



Parametric models

Parametric models (exponential, Weibull and log-logistic) tested show some misfit in the diagnostic plots, with the log-logistic function providing the best fit. When baseline covariates are included additively into this model (age and cognitive score), the model diagnostics improved (Figures 5 and 6)

Figure 6: Martingale residuals vs baseline cognitive score in base model (left) and covariate model (right)



References

1. Friberg LE, de Greef R, Kerbusch T and Karlsson MO. Modeling and simulation of the time course of asenapine exposure response and dropout patterns in acute schizophrenia. *Clin Pharmacol Ther.* 2009; 86(1):84-91
2. Collett D. (1994) Modelling survival data in medical research

Poster presented at Population Approach Group Europe (PAGE), June 11 – June 14, 2019, Stockholm, Sweden.