Online supplement

Joint longitudinal-survival model

Outcome data for each patient were of two types: a time series of renal function measures (eGFR) and the time to the initiation of renal replacement therapy. Both are measures of a single underlying process, the progressive loss of renal function until end stage renal disease (ESRD) is reached. Variation among individuals in the rate of loss produces variation in time to ESRD. While the two outcomes can be analyzed independently, neither analysis uses all of the data. A regression model for the longitudinal eGFR measurements, such as a linear mixed effects model (LMM), treats the interruption of eGFR measurements by the onset of ESRD as missing data rather than an outcome. This biases trajectory estimates because it violates the assumption of missingness at random (MAR). A survival analysis for the time to ESRD, such as a Cox proportional hazard model, ignores all the information on eGFR trajectories of patients who do not reach ESRD during follow-up observation.

A joint longitudinal-survival model uses both sources of information about the underlying process. The key advantages of this approach over separate analyses of the two outcome measures are bias reduction and gain in precision of estimates. The random effects formulation in the model takes the correlation of the two outcome measures into account.

We use the implementation of a family of joint longitudinal-survival models for SAS NLMIXED procedure, tailored for use with eGFR data and time to ESRD.

For the purpose of analysis, we specified the model as follows:

$$Y_{ij} = X_{ij}^{(l)}\beta^{(l)} + Z_{ij}b_i + e_{ij}$$

$$T_i = e^{X_i^{(s)}\beta^{(s)} + r b_i + \delta \varepsilon_i}$$

$$b_i \sim N(0, \Sigma)$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$\varepsilon_i \sim N(0, 1),$$

where Y_{ij} represents j^{th} longitudinal observation in i^{th} individual, T_i is subject's, possibly censored time to event, $X_{ij}^{(l)}$ and $X_i^{(s)}$ are vectors of covariates in longitudinal and survival processes, respectively, with associated vectors of fixed effects $\beta^{(l)}$ and $\beta^{(s)}$, Z_{ij} is a design

matrix formed by subset of columns of $X_{ij}^{(l)}$, b_i are random effects in longitudinal process with covariance matrix Σ, r are proportionality (scaling) parameters corresponding to elements of b_i, δ is scale parameter in survival sub-model, σ^2 is residual variance in longitudinal sub-model and e_{ij}, ε_i are residuals, conditionally independent of b_i, in longitudinal and survival sub-models, respectively.

The joint model assumes that subject-specific effects, such as random intercepts, slopes and quadratic terms (individual deviations from fixed effect predictions of baseline eGFR and rates of eGFR decline described by linear and quadratic function of time) are proportional to three components of the frailty term in survival (individual deviations of log-time to ESRD from the fixed effect-predicted log-time). Therefore an individual whose baseline eGFR is low, his renal function decline is fast or it accelerates will exhibit a short time to ESRD. On the other hand a patient whose baseline eGFR is high or eGFR decline is slow, or the decline decelerates will exhibit a long survival time. In other words, the joint model postulates that random effects describing a given subject's propensity to an outlying eGFR profile propagate to having an outlying value of frailty in the hazard model. The model was fitted with a marginal maximum likelihood method using adaptive Gaussian quadrature for approximation of the integral over the random effects and trust region optimization technique, replaced by Newton-Raphson technique, when non-convergence was detected.

Based on the Akaike information criterion, we chose a log-normal probability distribution for the observed time to ESRD over an exponential or Weibull parametric models. We performed residual diagnostics of the log-normal model using deviance residual plots against rank of survival time, and we also plotted conditional residuals and scaled marginal residuals from the longitudinal model against observation time.