

Variable selection (Cox regression model)

Independent variables analyzed

Age, sex, IPF (idiopathic pulmonary fibrosis), rheumatic disease (RD), cancer, tobacco exposure, symptoms at diagnosis, T/S ratio

Interactions: age \times T/S ratio, IPF \times T/S ratio, RD \times T/S ratio, cancer \times T/S ratio

Initial selection of independent variables
<p>Univariate hazard regression ($p < 0.3$)</p> <p>*Age, sex, *IPF, *RD, *cancer, tobacco exposure, symptoms at diagnosis, T/S ratio, * age \times T/S ratio, *IPF \times T/S ratio, *RD \times T/S ratio, cancer \times T/S ratio</p>
Refinement of initial selection
<p>1. Backward Stepwise selection. p-value for exclusion ($p \geq 0.1$)</p> <p>Selected variables: IPF, RD, age \times T/S ratio, RD \times T/S ratio,</p> <p>The initial model contemplates univariate selected variables $p = 0.9457 \geq 0.1000$, removing IPF \times T/S ratio $p = 0.9286 \geq 0.1000$, removing Cancer $p = 0.8516 \geq 0.1000$, removing Age $p = 0.1562 \geq 0.1000$, removing T/S ratio</p> <p>2. Forward Stepwise selection. p-value for inclusion ($p < 0.05$)</p> <p>Selected variables: IPF, RD</p> <p>The initial model is empty $p = 0.0352 < 0.0500$, adding IPF $p = 0.0367 < 0.0500$, adding RD</p> <p>3. Selection of the best equation by AIC/BIC and Harrell's C criteria</p> <p>Selected variables according to best AIC/BIC: age, IPF, RD, T/S ratio, RD \times T/S ratio</p> <p>Harrell's C: 0.814; AIC: 81.4; BIC: 91.7 (better AIC)</p> <p>Selected variables according to best Harrell's C: age, IPF, cancer, RD, T/S ratio, age \times T/S ratio</p> <p>Harrell's C: 0.825; AIC: 88.3; BIC: 100.6 (better Harrell's C)</p>

Variables selected in common: age, IPF, RD, RD \times T/S ratio

Diagnosis of hazard Cox model assumptions.

- Colinearity. Variance inflation factor (VIF) must be < 10 . Mean VIF: 4.87
- Presence of hazard proportionality (absence of relation between Schoenfeld residuals and survival time and absence of time varying covariates (tvc)). Non-significance means that null hypothesis H_0 for proportional hazards cannot be rejected.

- Test of proportional hazards assumption based on Schoenfeld residuals.

Global test. $p = 0.912$

rho Age. $p = 0.678$

rho IPF. $p = 0.830$

rho RD. $p = 0.325$

rho T/S ratio. $p = 0.530$

rho T/S ratio \times RD. $p = 0.277$

- Test of interaction terms between predictors and survival time.

Global test (*chunk test*). $p = 0.632$

tvc Age. $p = 0.711$

tvc IPF. $p = 1.00$

tvc RD. $p = 0.136$

tvc T/S ratio. $p = 0.466$

tvc T/S ratio \times RD. $p = 0.118$

- Test of interaction terms between predictors and logarithmic transformation of time.

Global test (*chunk test*). $p = 0.536$

tvc Age. $p = 0.545$

tvc IPF. $p = 0.534$

tvc RD. $p = 0.121$

tvc T/S ratio. $p = 0.556$

tvc T/S ratio \times RD. $p = 0.086$

- Presence of log-linear relationship between instant rate of incidence and predictors. Non-significance supports the hypothesis of linearity.

Predictor linear square analysis $p = 0.241$

- Influent observations (likelihood displacement).

2 subject detected as influents. No errors in transcription or measures.

After deleting these patients from dataset, there were no changes in significance of predictors, except for age which turned out to be significant (HR: 1.20; 95% CI: 1.04 to 1.38; $p = 0.013$).