Predict kidney

The Cox proportional hazard model is based on assuming a survival function $S(t)$ (the probability of surviving beyond t) of the form

$$S(t) = S_0(t)^{e^{\sum \beta_i}},$$

where $S_0(t)$ is the survival function of a 'baseline' patient (with all factors at 0), and β_i is the coefficient (log hazard ratio) associated with the ith factor.

Thus

$$-\log S(t) = -\log S_0(t) \times e^{\sum \beta_i},$$

and

$$\log[-\log S(t)] = \log[-\log S_0(t)] + \sum \beta_i.$$

Leibovich et al round $4 \times \sum \beta_i / 2.014$ to a score s, where 2.014 is the coefficient for 'pT3b, pT3c, and pT4'. Therefore $\sum \beta_i / \approx s/2$.

Therefore, if $S_s(t)$ is the survival function for score s, we can reverse engineer to get the baseline survival by

$$\log[-\log S_0(t)] \approx \log[-\log S_s(t)] - s/2.$$

The survival functions for different scores are given in Table 5, although these appear to have been calculated independently for each group as Kaplan-Meier curves, rather than coming from the Cox model. Nevertheless we should be able to use the approximation above to estimate the baseline survival (see attached spreadsheet).

The estimates coming from scores 4 to 7 seem reasonably consistent, and averaging over these gives an estimated baseline survival function of $\hat{S}_0(t) = 0.992$, 0.981, 0.976, 0.972, 0.967 for 1,2,3,7,10 years.

Reverting to the original Cox model, the survival for anyone with coefficients β_i can then be approximated by the original formula

$$S(t) = \hat{S}_0(t)^{e^{\sum \beta_i}}.$$