Time dependent covariates and absolute risk

Terry Therneau

19 Sept 2023

1 Introduction

A common question, and pitfall, is how to create survival curves from a Cox model with time-dependent covariates. A survival curve is a statement, at some given point in time, of predicted future survival given known covariates *at that time*. When the model includes time-dependent covariates, such a prediction requires, as well, joint prediction of future covariate paths for any time-dependent covariates. Standard software does not provided this joint prediction, instead there is wide opportunity to produce incorrect or misleading results. (Note that this is a different question than time-dependent coefficients, which appear in a discussion of proportional hazards.) As an illustration, we reprise an example from Section 10.2.4 of [**?**].

This study recuited subjects with primary biliary cholangitis to a placebo controlled trial of D-penicillamine. PBC is a chronic condition that at the time had no effective therapy. During the recruitment period 424 patients met the eligivility criteria; 312 agreed to participate fully and another 108 to initial laboratory measurements and long term follow-up. Sequential laboratory data and further follow-up is available on the 312 enrollees. The survival package data set `pbc` contains baseline values and survival for all 418 and `pbcseq` the sequential lab values. As background, table 1 shows the coefficients for each of the 5 variables used in the risk score model of Dickson et al (doi: 10.1002/hep.1840100102), 3 models fit to the extended data, and the model concordance. The models using baseline data for all 418 and the 312 in the study have essentially the same predictive power, but the time-dependent covariate models are clearly

	All 424 Time fixed	312 Time fixed	312 Time dependent	
age	0.57	0.56	0.65	
edema	0.90	0.60	0.76	0.52
log(bili)	1.25	1.24	1.55	1.62
albumin	0.37	0.34	0.67	0.63
log(protime)	0.25	0.45	0.30	0.36
concordance	0.83	0.84	0.91	0.90

Table 1: Standardized coefficients for the original PBC model using all 424 subjects, and using the 312 study subjects with either time fixed or time dependent predictors on entry scale, and then time-dependent on age scale. These represent the effect of a 1 unit change in edema, and a change from the 25th to 75th percentile for the other 4.

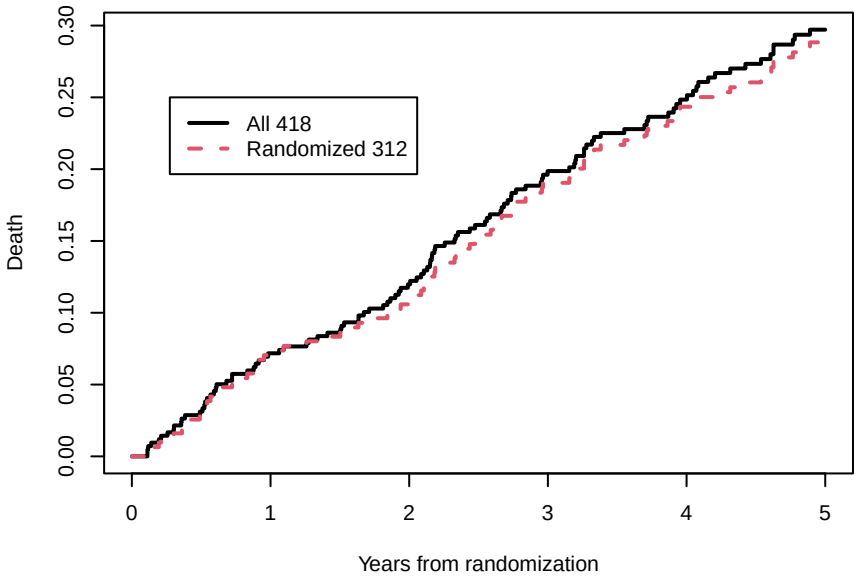

Figure 1: Comparison of the 312 randomized to the full cohort.

stronger. Bilirubin is the most important predictor: a subject at the 75th percentile has exp(1.2) $\approx$ 3.3 fold hazard as someone at the 25th percentile of bilirubin.

As an aside, the coefficients for enrollment time, in the age scale model, are not significant but also not trivial (estimated 1.3 increase). But the first 5 years' survival for the study and expanded cohort essentially overlap.

```
> print(pcheck, digits=2)
Call:
coxph(formula = Surv(age1, age2, death) ~ edema + log(bili) +
    albumin + log(protime) + year, data = pdata)

              coef exp(coef) se(coef)    z      p
edema         0.54      1.72     0.24  2.2   0.03
log(bili)     1.10      2.99     0.11  9.6 <2e-16
albumin      -1.35      0.26     0.21 -6.4  2e-10
log(protime)  3.41     30.35     0.69  5.0  7e-07
year0-1       0.29      1.34     0.28  1.0   0.30
year1-3       0.34      1.41     0.22  1.5   0.13

Likelihood ratio test=429  on 6 df, p=<2e-16
n= 2472, number of events= 140
```

Figure 3 helps explain the superiority of the time-dependent model. To paraphase one of the MD investigators, the liver has a moderate amount of excess capacity. As the disease's

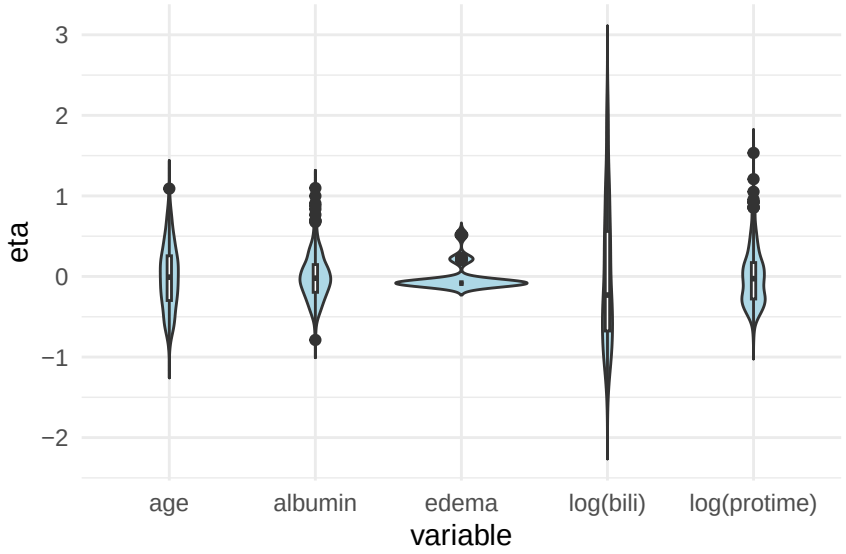

Figure 2: Relative effect of each covariate on the risk score η.

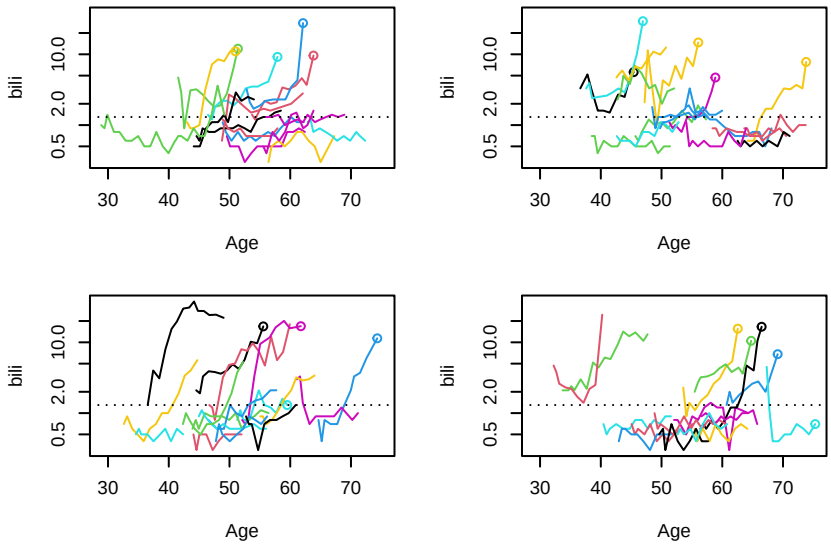

Figure 3: Bilirubin trajectories for the 53 subjects with 10 or more laboratory measurements, separated into 4 panels to decrease overlap. A bilirubin of ≤ 1 is normal, and above 1.3 begins to be a cause for concern (the horizontal dotted line). Deaths are marked with a circle.

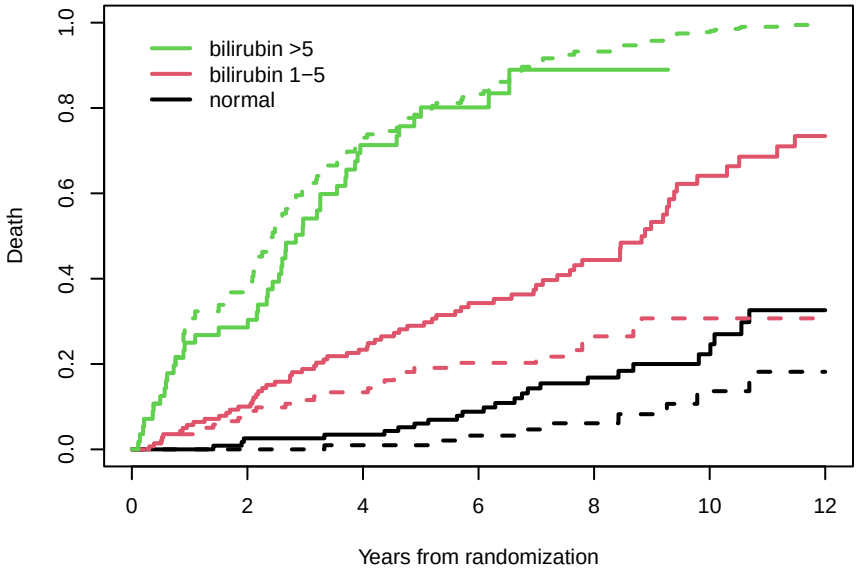

Figure 4: Survival from baseline for the 3 bilirubin groups. Solid lines are the Kaplan-Meier, dashed are the "extended Kaplan-Meier".

inflammatory process continues, it steadily converts functioning tissue to scar, and liver function tests slowly rise to a point but then rapidly increase when that excess capacity is exhausted. At the start of the trial subjects ranged from early to late disease; as time goes on we see that many of them experience the bilirubin acceleration.

1.1 KM curves

For illustration we will divide the subjects into 3 groups: normal bilirubin of ≤ 1, 1–5 and 5+. Figure 4 shows the standard Kaplan-Meier for each of the 3 subgroups as a solid line, divided by bilirubin level at baseline. The dashed lines are the "extended Kaplan-Meier" recommended by Snapinn et al [**?**]. These curves make direct use of the time-dependent data set: at each time point the increment to bili>5 curve is based on the set of subjects currently at risk and *currently* having a bilirubin level that is > 5. In the usual KM the increment is based on the initial bilirubin. The impact on counts is shown in table 2. Notice that nearly all the death (100/140) are now attributed to the high bilirubin group.

How do we interpret these curves? The simple KM is simple: each curve is an estimate of the future survival, from randomization, for a set of subjects in the given state at randomization. No use is made of the follow-up lab values. For the extended KM, the argument is made that the normal bilirubin curve estimates the survival of subjects who start and remain in that state, i.e., their bilirubin never rises above 1, and likewise that curve 2 represents subjects whose bilirubin remains between 1 and 5. A more cautionary note is provided in [**?**], who looks at the risk sets more carefully from a causal models perspective; who finds that underlying premise that those currently in group 1–3 represent subjects who are always in that state requires additional strong

	Number at risk				Total
	0	4y	8y	12y	Deaths
Usual KM					
Normal	116	112	58	14	25
1–5	140	99	45	11	70
5	56	14	1	0	45
Extended KM					
Normal	116	100	51	10	9
1–5	140	92	42	12	31
5	56	33	11	3	100

Table 2: Number at risk a 0, 4, 8, and 12 years along with the total number of deaths for each group, for the two estimators.

assumptions. Our view is more simple: even if the curve can estimate what it claims to estimate, of what use is it? In this disease the liver status will invariably seriously decline over time; the estimator has created curves for someone who does not exist.

An alternative that does make use of the evolving laboratory data, but also estimates a quantity of direct interest, is a multi-state model shown in figure 5. Figure 6 shows the predictions for each starting state along with the simple KM. The multi-state curve for the 'normal bilirubin state has an increment, at each death time, of

$$\sum_i P(s(t) = i | s(0) = 1) P(\text{death} | s(t) = i)$$

where $s(t)$ is the state. We speculate that the increase in death rate for the 'normal' curve is a reflection of reclassification as subjects go between the first and second state.

1.2 Hazard models

We can repeat the same exercise with predicted curves from a Cox model. Figure 7 contains the overall KM as a reference, along with the predicted curve based on the fit using baseline values, and the fit using time-dependent covariaties. In both cases these are for a subject with average covariate values. The simple curve (dotted) is quite a bit below the KM, which is a consequence of the fact that $E(f(X)) \neq f(E(X))$ for any non-linear function f. The predicted survival curve from a Cox model is a quite non-linear function of the linear predictor η, we have drawn the curve for an average η. The marginal prediction from the model is an average of all $n = 312$ predicted curves from the model, one for each subjects, and it is very close to the KM. (Not exactly identical due to differential follow-up and the PH assumption.)

The prediction from the time-dependent model suffers from the same issue as the extendend KM: it is attempting to predict the survival of a subject who starts with these average covariate values, and then never changes. The computational mechanism is different than the EKM estimate: all observations participate in creating a single baseline hazard estimate, and this is then used to predict the curve for a subject with constant covariates. Whether said computation is justifiable or not, in our opinion the result is simply not interesting.

A possible solution would be to create a marginal curve. The software allows for specification of a time-dependent covariate path, and will then produce the curve corresponding to that

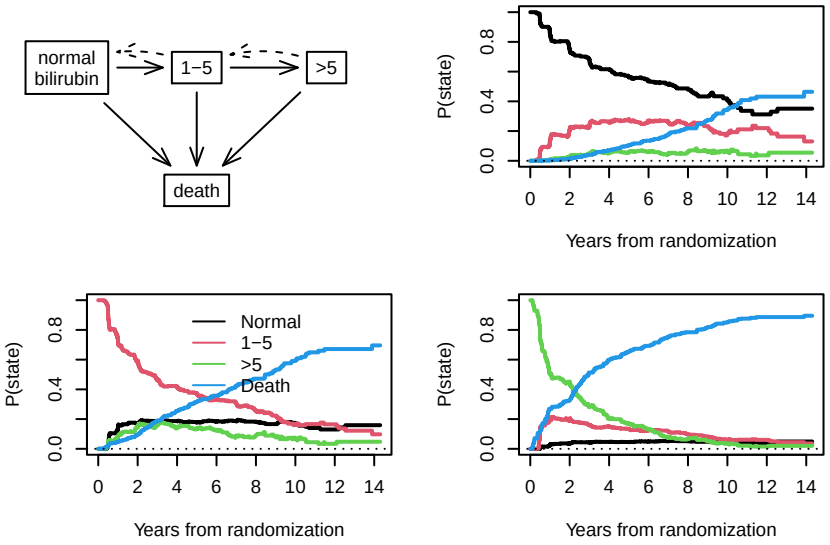

Figure 5: Potential state space for the PBC data, along with Aalen-Johansen estimates assuming that everyone starts in the normal state (upper right), the bilirubin 1–5 state (lower left), or bilirubin > 5 state (lower right).

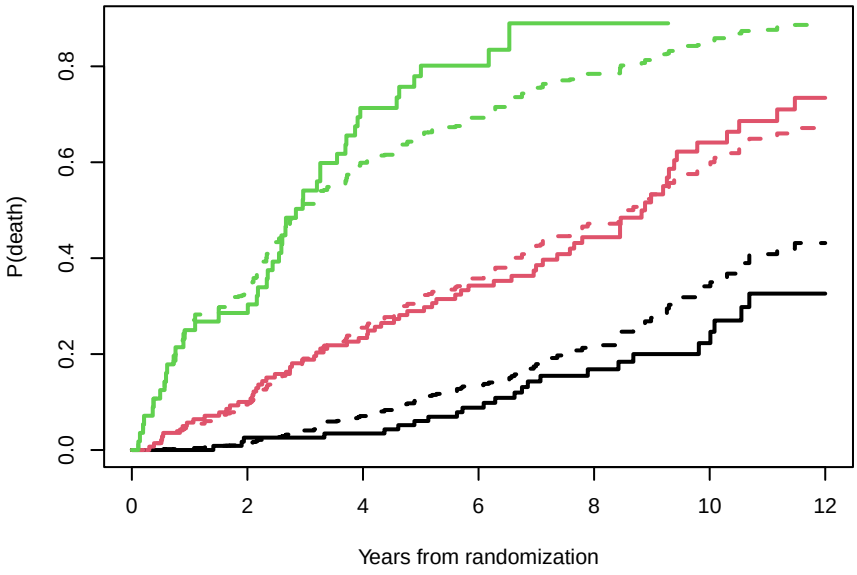

Figure 6: Standard Kaplan-Meier (solid) along with multi-state estimates of survival for the PBC data.

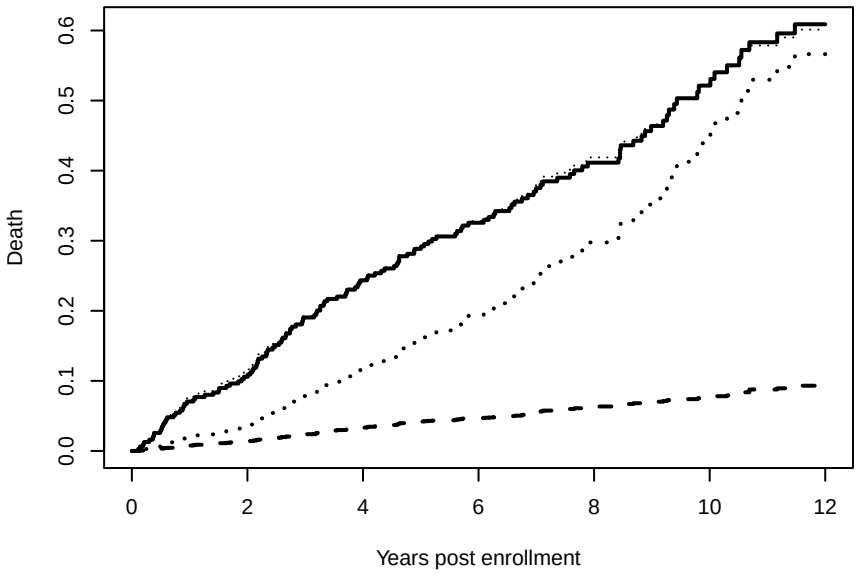

Figure 7: Overall survival for the PBC dataset (dotted line), along with predicted survival from the time-fixed (solid) and time-dependent Cox models for a subject with median covariate values at enrollment: age, bilirubin, edema, albumin and prothrombin times of 50, 1.35, absent, 3.5 and 10.6, respectively. The marginal curve is shown as a dotted line, and is nearly coincident with the KM.

path. Compute the curves for all n patients and average them. The problem with this idea is differential follow-up. Subject 1, for instance, died at day 400. Unless we are willing to project some hypothetical future covariate path for the observation, the predicted curve also stops after just over 1 year. Taking a simple average, i.e., marginal(t) = mean of all curves defined at t, will certainly be discontinuous and may not even be monotone.

Multistate models again offer a potential solution to the issue. Consider the following model:

```
> test <- survcheck(Surv(day1, day2, bstate) ~1, pdata, id=id, istate= bili3)
> test$transitions
        to
from     normal 1-5  >5 death (censored)
  normal      0  94   0     9         77
  1-5        64   0 101    31         65
  >5          0  27   0   100         30
  death       0   0   0     0          0
>
> mfit1 <- coxph(list(Surv(day1, day2, bstate) ~ 1,
                      c(1:3):"death" ~ age + edema + albumin +
                          log(protime) / common + shared),
                data= pdata,  id=id, istate= bili3)
> print(mfit1, digits=2)
Call:
coxph(formula = list(Surv(day1, day2, bstate) ~ 1, c(1:3):"death" ~
    age + edema + albumin + log(protime)/common + shared), data = pdata,
    id = id, istate = bili3)


1:4               coef exp(coef) se(coef) robust se    z     p
  age           0.0332    1.0338   0.0086    0.0112  3.0 0.003
  edema         0.9903    2.6920   0.2329    0.2201  4.5 7e-06
  albumin      -1.3524    0.2586   0.2034    0.2265 -6.0 2e-09
  log(protime)  3.7080   40.7720   0.5784    0.6060  6.1 9e-10


2:4               coef exp(coef) se(coef) robust se    z     p
  age           0.0332    1.0338   0.0086    0.0112  3.0 0.003
  edema         0.9903    2.6920   0.2329    0.2201  4.5 7e-06
  albumin      -1.3524    0.2586   0.2034    0.2265 -6.0 2e-09
  log(protime)  3.7080   40.7720   0.5784    0.6060  6.1 9e-10
  ph(1:4)       0.5797    1.7856   0.3877    0.3787  1.5 0.126


3:4               coef exp(coef) se(coef) robust se    z     p
  age           0.0332    1.0338   0.0086    0.0112  3.0 0.003
  edema         0.9903    2.6920   0.2329    0.2201  4.5 7e-06
  albumin      -1.3524    0.2586   0.2034    0.2265 -6.0 2e-09
```

```
  log(protime)  3.7080   40.7720   0.5784    0.6060  6.1 9e-10
  ph(1:4)       2.1542    8.6211   0.3720    0.3420  6.3 3e-10

 States:  1= normal, 2= 1-5, 3= >5, 4= death

Likelihood ratio test=440  on 6 df, p=<2e-16
n= 2472, number of events= 140
>
> msurv1 <- survfit(mfit1,newdata=dummy, p0=c(1,0,0,0))
>
> test <- coxph(Surv(day1, day2, death) ~ age + edema + albumin + log(protime)
               + bili3, data= pdata)
```

This is a more complex model. I have forced no covariates for the transitions between bilirubin states. Bilirubin itself should not be a preditor of those, due to edge effects if nothing else, e.g., a subject with bili of 4.9 is more likely to transition to the 5+ state than someone with a bilirubin of 3 simply due to measurement variability. Proportional hazards has been assumed for the 3 transitions to death, resulting in scale factors of 1, 1.7, and 2.2. Though not printed above, the coefficients and log likelihood of the test fit are identical to mfit1. The advantage of mfit comes when we ask for a predicted survival curve: the bilirubin level is given as an intial state which then evolves over time, rather than a fixed numerical value which does not. The other three covariates still appear as time fixed variables, so this solution is incomplete, but it gives the flavor. A plot of the death state for msurv1 corresponds to death for a hypothetical subject whose bilrubin can evolve but other variables remain fixed (perhaps even more uninterpreatable than the EKM). Bilirubin is by far the strongest predictor, however, and it is not surprising that the curve lies between the dashed and dotted lines of figure 7 (not shown).

1.3 Using age scale