Time-dependent covariates and absolute risk

Terry Therneau

18 Sept 2024

```
## Loading required package:  Matrix
##
## Attaching package:  'expm'
## The following object is masked from 'package:Matrix':
##
##     expm
```

1 Introduction

This is a slightly modified version of a draft chapter for my book with Cynthia Crowson and Elizabeth Atkinson. A *draft* version. There is some text in "comment" blocks, which should show up in blue on a screen; these are asides between the authors and the authors/readers. This is a very new idea, it seems to work, but there are a lot of unknown details.

Time-dependent covariates represent a conundrum. They have proven to have great practical utility in proportional hazard models, however creating survival curves with respect to a time-dependent covariate has proven to be very problematic. We have argued in this book that for interpreting a multi-state hazard model (MSH), that *both* hazards and absolute risks are necessary. How can we resolve this tension and make use of time-dependent covariates in multistate models?

For the hazard portion of the multi-state hazard (MSH) model, which is a model for the instantaneous hazard at each time t, using all the available information up to time t makes complete sense. When we see our physician for care of a chronic disease, we want them to use all the information in our medical history up to that point of care, not limit their assessment only to information available at time 0, the date of disease onset.

The problem comes when creating estimates of absolute risk from such a model. A survival curve is by definition a predicted future trajectory, based on information known at the present time. Use of time-dependent covariates in this exercise risks the logical fallacy of using the future to predict the future. Kalbfleish and Prentice argued that although a time-dependent hazard function can be defined and then simply integrated, that “this hazard bears no relationship to a survivor function” [2] (Section 5.3.2). Our own 2000 monograph [5] took a somewhat less rigid but still very pessimistic view.

In Section 2 we will review several approaches to survival using time-dependent covariates, both successful and unsuccessful, using both the familiar PBC trial data and a simulated dataset as examples. In Section 3 this will be applied to a study focusing on risk factors for dementia, where time-dependent covariates, multistate hazard models, and predicted absolute risk all play an important role in the analysis.

As a starting point we want to distinguish between three issues

1. External time-dependent covariates: these are values that may change over time, but are known independent of an individual subject. An example might be daily weather in a study of the risk of hip fracture; e.g., falls might be more likely when it is icy. Another would be a subject’s current age, which can be perfectly predicted from enrollment age.

2. Internal time-dependent covariates: values which require ongoing patient measurement. Examples are blood pressure, laboratory tests, or cumulative comorbidities.

3. Time-dependent coefficients: the estimated coefficient effect $\beta(t)$ varies over time, but not the covariate.

The issues discussed here arise in case 2.

A fitted multistate hazard model with time-fixed covariates provides a predicted relative hazard for any specified covariate vector, and also predicted absolute risks for that covariate vector. These absolute risks include the probability in state $p(t)$, the expected sojourn time in each state, and the expected number of visits to each state. Population marginal means, Chapter **??**, can used to provide absolute risks indexed by a single predictor. The approaches found in that Chapter also work well for either case 1 or case 3 just above. There may be a bit more nuisance in the computation of curves, e.g., deciding on a “interesting” weather pattern for which we would like to compute the cumulative number of fractures, but there is no conceptual change in the computational process. Internal time-dependent covariates are the difficult case, with both computational and conceptual challenges. The latter, clear thinking about exactly what is to be estimated and its practical use, is actually the more challenging issue.

2 Existing Methods

2.1 Data

The PBC dataset includes 312 subjects recruited to two placebo controlled trials of D-penicillamine, along with 106 who met eligibility requirements who declined full participation but agreed to initial laboratory measurements and long term follow-up [1]. Sequential laboratory data and further follow-up is

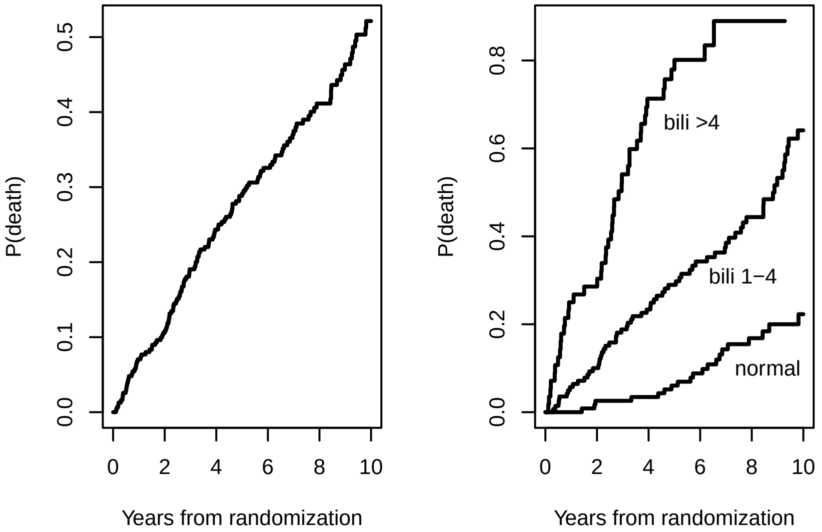

Figure 1: Left panel: overall survival of the 312 randomized participants. Right panel: Survival by enrollment bilirubin level.

	Hazard Ratio		
	Age10	log2(bili)	C
Time-fixed	1.6	2.1	0.81
Time-dependent	1.9	2.6	0.88

Table 1: Coefficients and concordance values (C) for Cox models using age and bilirubin. Age10 = age in decades.

available on the 312 trial enrollees, who form the basis for the examples here. The trial unfortunately showed that D-penicillamine had essentially no effect on survival, and the study data has since found its greatest use as a description of the natural history of the disease.

The overall Kaplan-Meier curve is shown in Figure 1. The curve is nearly linear with a loss of approximately 5% of the enrollees each year. The classic PBC risk score has 5 variables: log(bilirubin), age, log(albumin), log(prothrombin time) and edema score [1]. However, a model with only the first two of these variables accounts for 80% of the gain in log-likelihood, so for simplicity we will use only enrollment age and bilirubin in the examples of this section, bilirubin as either continuous or a simple categorical (≤ 1, 1–4, > 4). The upper limit of normal for bilirubin is 1.0. About 21% of the subjects have a bilirubin > 4 at enrollment, and 48% reach a level over 4 sometime during follow-up.

Table 1 shows results for proportional hazards models based on baseline (enrollment) bilirubin and using time-dependent bilirubin values, along with

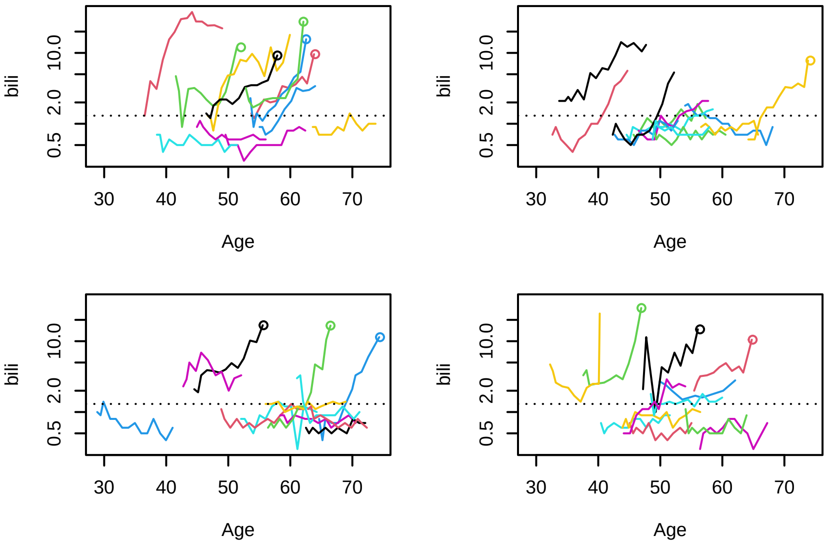

Figure 2: Bilirubin trajectories for the 48 subjects with 11 or more laboratory measurements, separated into 4 panels and shown on age scale to decrease overlap. A bilirubin of ≤ 1 is normal, and above 1.3 begins to be a cause for concern (the horizontal dotted line). Deaths are marked with a circle.

age in decades as a second covariate. For the time-fixed model each decade of age increases the hazard 1.6-fold and each doubling of bilirubin approximately doubles the hazard. The time-dependent model shows a stronger bilirubin effect and a greater predictive accuracy.

Figure 2 helps explain the superiority of the time-dependent model. To paraphrase the primary investigator, a normal liver has a moderate amount of excess capacity, but as the disease's inflammatory process continues, this slowly but steadily converts functioning tissue to scar (cirrhosis). 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 this turning point in bilirubin values. (Medically, every patient is guarrateed to experience such a rise if they live long enough, although cardiovascular events are also a leading cause of death.)

One concern for the time-dependent dataset is that an excessive bilirubin may prefigure death yet be unrelated to PBC, e.g., the hyperbilirubinemia that can occur during multiple organ failure. For this analysis we have chosen to ignore any laboratory measurements within 7 days of death, which comprise 15 of the available bilirubin values.

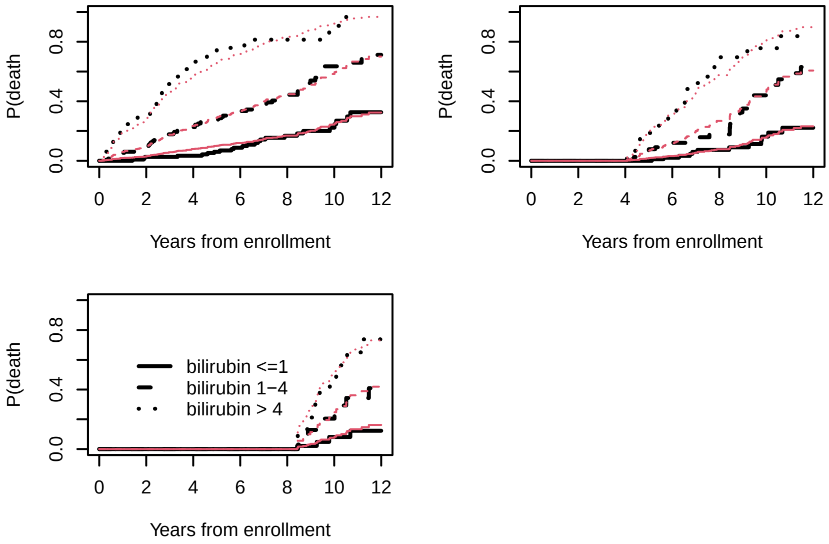

Figure 3: Landmark curves for the PBC study at 0, 4, and 8 years (black), along with predicted curves from landmark Cox models (red). Groups are based on bilirubin levels at 0, 4, and 8 years, respectively.

	Age 10	bilirubin 1–4	bilirubin ¿ 4	n1	n2	n3
Baseline	0.4	1.2	2.4	116	131	65
Year 4	0.6	1.4	2.5			
Year 8	0.7	1.3	2.4			

Table 2: Coefficients for the landmark Cox models, along with the number of subjects in each of the bilirubin groups.

2.2 Conservative approach

The conservative approach is to draw only time-fixed curves, starting either at enrollment or at some fixed time τ after enrollment. The latter is referred to as a landmark analysis.

Figure 3 shows the result of landmark Kaplan-Meier curves at 0, 4, and 8 years after enrollment. The first of these is the usual KM, with subjects divided into groups using baseline bilirubin. The second uses the bilirubin values at 4 years, for all those alive at 4, and the third values at 8 years. Overlaid on each are the marginal predictions from a Cox model using age and bilirubin group as covariates, and based on the same data. TT: Add the KM and Cox CI intervals? BA: The Cox will be smaller. I don't think it is necessary here.

The problem with the landmark curves is that they are correct but may be irrelevant. For a patient looking ahead at time 0 (enrollment), the curves in the upper left are the relevant ones for future planning. The 4 year landmark curves

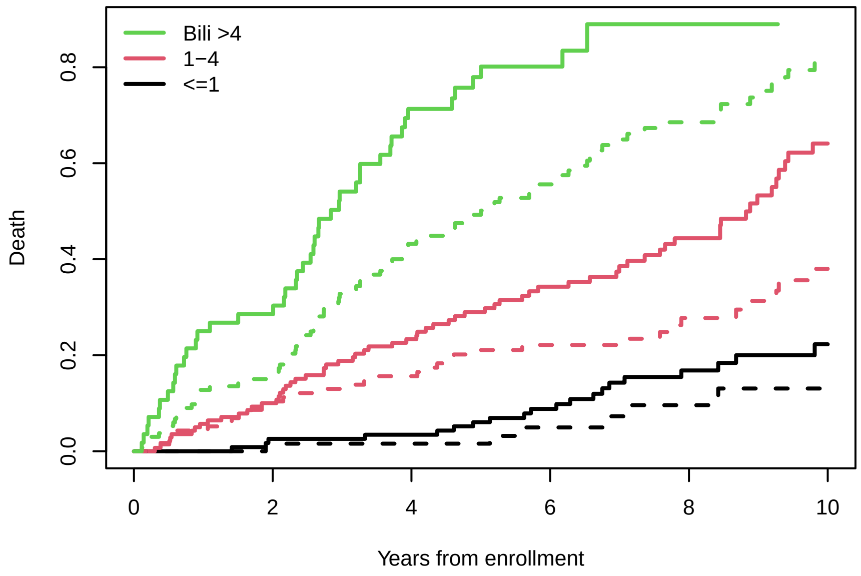

Figure 4: Survival curves categorized by bilirubin level at baseline (solid) or by maximum observed bilirubin over follow-up (dashed).

will become relevant at the patient's four year visit, should they live that long, but until then are simply hypothetical.

2.3 Ineffective methods

Immortal time bias

Possibly the worst thing that can be done is to treat a time-dependent covariate as though it were fixed, using information from the future as that fixed value. This is an example of immortal time bias, discussed more fully in Section **??**. As an example using the PBC data, first create a variable which categorizes each subject by their maximum observed bilirubin, and then use that as a baseline covariate. Figure 4 compares this to a categorization by baseline bilirubin values at enrollment. All of the dashed curves are biased downwards. The "bilirubin ever > 4" group, for instance, includes 56 subjects with an enrollment bilirubin over 4 along with 72 others whose observed bilirubin exceeded 4 at some later time. To get into the 72 you need to live long enough for bilirubin to increase all the way to 4. Thus the "> 4" label systematically selects for those who live longer. The downward bias in the graph is a visual encoding of the maxim that "those who live longer have a better survival", which is not a particularly deep insight. The dashed curves are not interpretable in any clinically meaningful way.

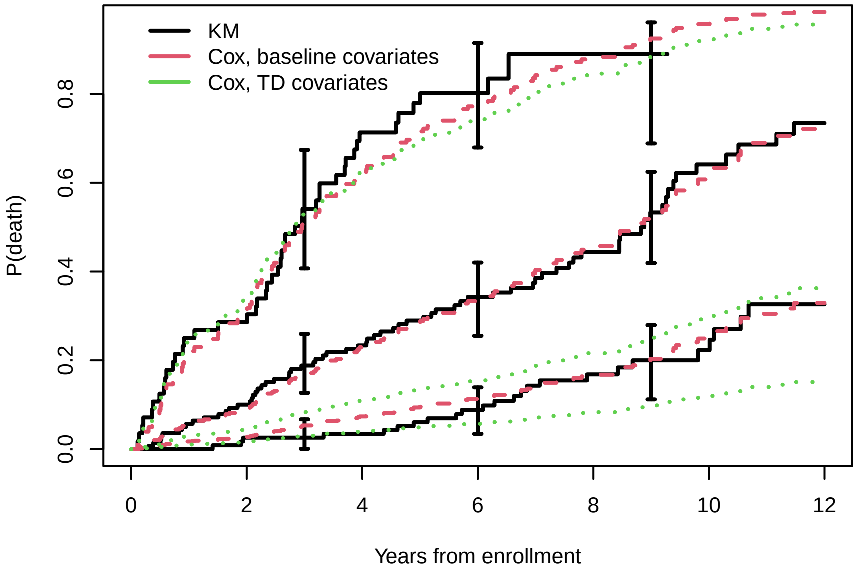

Figure 5: Predicted survival curves for the PBC data, using the KM based on baseline values (solid), predictions from a Cox model fit using baseline covariates of age and bilirubin group, marginal over age, and and marginal prediction from a Cox model using time-dependent data.

Static referrent

A method in which the population is allowed to undergo dynamic change, i.e., time-dependent covariates, but the referent remains static has been known for a long while, and was seriously critiqued in [5]. To compute this, first fit a Cox model using time-dependent bilirubin. Table 1 show coefficients and C statistics for said time-dependent fit along with a fit based on baseline bilirubin. Each doubling of baseline bilirubin leads to a predicted 2.1 fold increase in the hazard for death, but 2.4 fold using the time-dependent form.

Given the advantages of the time-dependent Cox model, it is natural to want to use the results of that model for predicted survival. Figure 5 shows the Kaplan-Meier curve for the dataset as a whole, along with overall predictions based on Cox model fits using either the baseline or time-dependent covariates. In this case, we used models with the categorical predictor, and predictions for three hypothetical cohorts from the same bilirubin groups. The predictions are marginal with respect to enrollment age, as discussed in Chapter **??**. that is, obtain 312 predicted curves for a cohort who all start in the normal bilirubin group at the 312 observed ages, and average the curves. Repeat for the other two groups. The overlap with the KM and first prediction attests to an excellent fit of the Cox model to this particular dataset.

The predicted curves from the time-dependent model look very different for the normal bilirubin group, with a 12 year predicted death rate that is less than half of observed; for the 1–4 group the time-dependent curve is .58 of observed. Why is it so different? These are the predictions for three hypothetical cohorts of 312 subjects with the same enrollment age distribution as the sample, one with all subjects in the normal range, one with all subjects in the 1–4 category and one with all subjects in the > 4 category, but each a cohort whose bilirubin level *never rises* over time. Though mathematically correct, the lower and middle curves each describe a hypothetical patient population that, unfortunately for this particular disease, does not exist. We have mixed a time-dependent estimator with a static prediction.

An estimate known as the extended Kaplan-Meier (EKM) [4] aims at the same target, but uses a variation of the Kaplan-Meier curve. At each time point the EKM computes the update to the bilirubin 0–1 curve, with respect to death as the endpoint, using all subjects who are *currently* in the normal state, whereas the ordinary KM uses all those who were in the normal state at time 0. For the EKM, the authors argue that this "0-1" 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 a hypothetical cohort whose bilirubin remains between 1 and 4. A more cautionary note with respect to that argument is provided in [3], which looks at the risk sets more carefully from a causal models perspective. They find the underlying premise that those currently in each category represent subjects who are always in that state requires additional strong assumptions. Our view of the EKM is more harsh: even assuming that the curve can estimate what it claims to estimate, of what practical use is it? In this disease the liver status will invariably worsen over time. There is no such thing as a patient whose bilirubin always remains normal.

The Kaplan-Meier, EKM, and prediction from a time-dependent Cox model are shown in Figure 6. The EKM for bilirubin 0–1 and 1–4 tracks the predictions from the time-dependent Cox model, with the caveat of a larger standard error. The reason for the SE difference is that the Cox model is able to use all subjects to estimate all curves due to the PH assumption, while the EKM for the 0–1 group, for instance, very quickly reduces in sample size from 112 at 2 years to 78 at 6 and 27 at 10 years. The fact that the estimates themselves are very similar arises from the fact that they are targeting the same quantity, i.e., a group whose bilirubin stays frozen over time. The curves answer a question that nobody asked.

To do, add the CI bars, without the plot getting too busy. Note that the predicted curve at (mean age, mean log(bili)) has a very different shape, this is an example of why marginal curves are important. Do we want to note that here, or is it TMI?

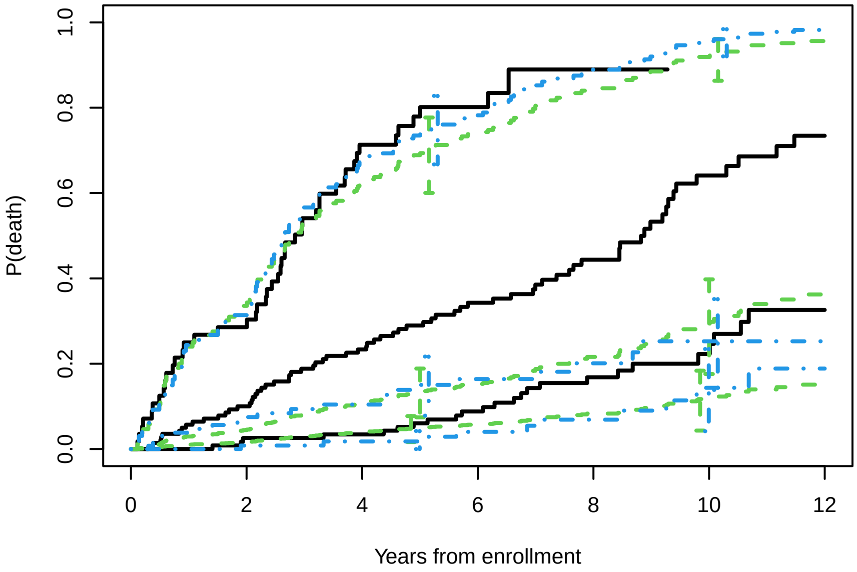

Figure 6: Kaplan-Meier curves for the PBC data, baseline bilirubin of 0-1 (normal), 1–4, and > 4, as solid lines; the extended KM (dashed) using time-dependent bilirubin, and predictions from a Cox model using time-dependent bilirubin (dotted).

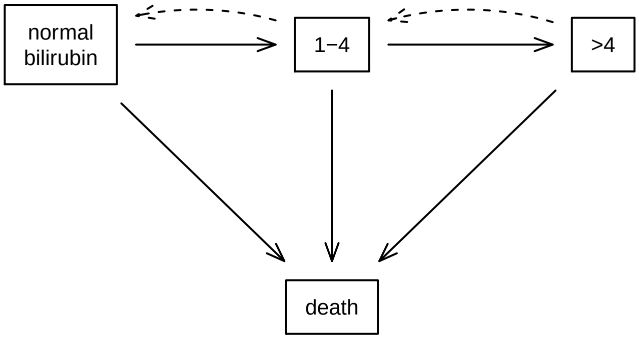

Figure 7: Expanded state space for the PBC data.

3 Prediction using a multistate model

3.1 Aalen-Johansen

The key philosophical issue with predicting outcome when there is a time dependent covariate is that we need to predict the future path of that covariate, in addition to predicting the outcome itself.

An alternative that explicitly makes use of the evolving laboratory data, but also estimates a quantity of direct interest, arises from the multistate model and Aalen-Johansen fits shown in the upper left of Figure 7, which treats both the bilirubin level and death as states. Recall the details of the Aalen-Johansen computation from Section **??**; it can be written as a product

$$p(t) = p(0) \prod_{s \le t} T(s)$$

Here $p(0)$ is the initial distribution of states. For the survial package this defaults to the observed distribution just before the first event, which in turn is most often a priori (1, 0, 0, ...), e.g., everyone starts in the 'enrollment' state. In this data set the default `p0` is (116, 140, 56, 0)/312.

For this 4 state model there is a 4x4 transition matrix $T(s)$ for each time at which a transition occurs; the T matrices are based on the observed data. The first T matrice in the data is $T(41)$, a death at day 41 of someone in the

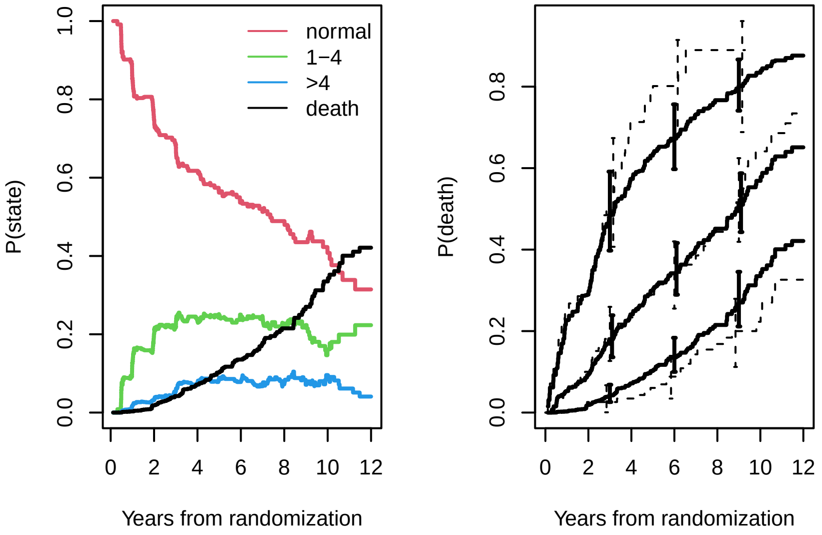

Figure 8: Left panel: Aalen-Johansen for the 4 state model of Figure 7 when all start in the normal state. Right panel: standard Kaplan-Meier (dashed light) along with multistate estimates (heavy line) of survival for those with an initial bilirubin of ≤ 1 (lower curves), 1–4 (middle) or > 4 (top).

bilirubin > 4 state and has the form

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 64/65 & 1/65 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We make use of the fact that in the survival package $p(0)$ can be set explicitly. The left panel of Figure 8 shows the results with $p(0) = (1, 0, 0, 0)$; this is the estimated probability in state curves for a hypothetical cohort of subjects who all start in the bilirubin=normal state. By about 10.5 years the estimated fraction who have died and who are still in the normal state are each about 35%. The right hand panel shows the simple KM using baseline data, and overlays 3 probability of death curves, one from each of the 3 AG fits with $p(0) = (1, 0, 0, 0)$, $p(0) = (0, 1, 0, 0)$ and $p(0) = (0, 0, 1, 0)$. The multistate curves closely track the KM but with a lower variance. Note that the T matrices remain identical for all the estimates, only $p0$ changes.

Let $Y_{ij}(t)$ be 1 if subject i is in state j and at risk for a transition at time t, $dN_{jk}(t)$ number of j:k transitions at time t, d the death state, and consider the increment from state 1 to death at each time t. For the simple KM estimate the increment is Equation (1) below, based on the death rate of those who started in state 1 and are still in state 1. The AJ increment at t is in Equation (2); it is a weighted sum of all three death rates at t, 1:d, 2:d, 3:d, using all subjects at risk at t, weighted by the probability that someone who started in state 1 will currently be in state 1, 2, or 3.

$$dKM = p_1(t-)\frac{dN_{1d}}{\sum_i Y_{i.}(t)Y_{i1}(0)} \tag{1}$$

$$dAJ = \sum_{k=1}^{3} p_k(t-|Y_{i1}(0)=1)\frac{dN_{kd}}{\sum_i Y_{ik}(t)} \tag{2}$$

The reason for the smaller variance is that the AJ estimate of death uses more of the subjects in the sample, effectively making use of a larger n. The variance is based on the infinitesimal jackknife (details in the formula appendix); the KM has 116 with non-zero influence and the AJ 295. There is of course no free lunch — the AJ estimate depends critically on the Markov assumption that the death rate depends only on the current state, and not the path taken to that state. Looking again at the figure, the multistate curves for the lowest and highest bilirubin groups hint at a systematic shift from the usual KM, though this is not statistically significant. This may perhaps reflect partial failure of the Markov assumption, e.g., a recent transfer from state 2 (bili 1–4) to state 3 likely has a lower average bilirubin than someone in state 3 from the start.

Perhaps in the companion show how to get the sd for comparing them.

Perhaps in the companion, show the transition matrix for the data. There is some back and forth between states. BA: I'm not sure this is helpful

Below is a repeat, don't know which is better. Another way to view the AJ estimate of death is to look at it as a weighted sum. Consider two estimates of the death rate at any given time t: $\lambda_k(t)$ = the death rate among all those who began in state k and are currently alive, and $\lambda^*_k(t)$ = the death rate among all those who are currently in state k and alive. Formally

$$\begin{aligned}
\lambda_k(t) &= \frac{\sum_{i=1}^n Y_{i1}(0)dN_i(t)}{\sum_{i=1}^n Y_{i1}(0)Y_i(t)} \\
\lambda^*_k(t) &= \frac{\sum_{i=1}^n Y_{i1}(t)dN_i(\tau)}{\sum_{i=1}^n Y_{i1}(t)}
\end{aligned}$$

where $Y_{ik}(t) = 1$ if observation i is alive and at risk in state k at time t.

The ordinary KM is based on λ, the EKM on λ^* and the multistate death estimate on a weighted sum

$$\lambda^m_k(t) = \sum_{j=1}^{3} P(s(t)=j|s(0)=k)\lambda^*_j(t) \tag{3}$$

The first term is the probability of currently being in bilirubin state j given they started in state k The second term uses everyone currently in state j to estimate the death probability from that state.

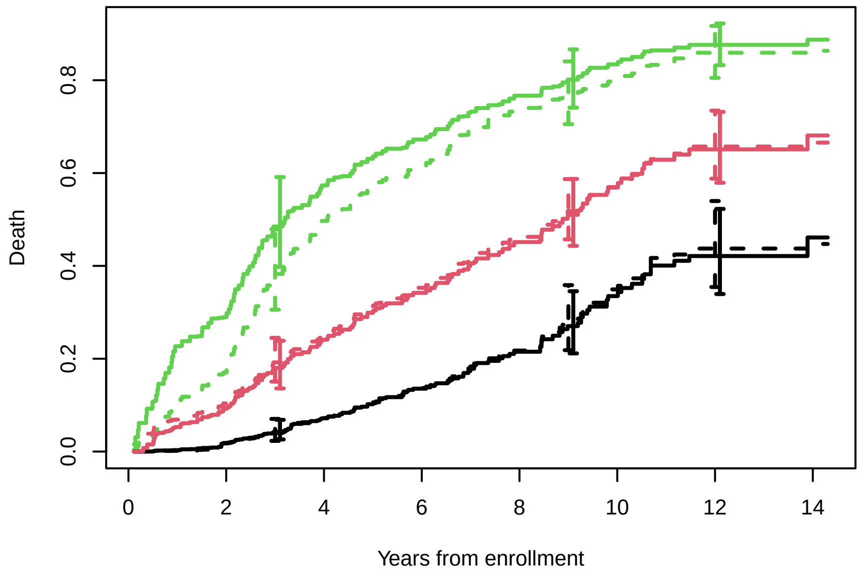

Figure 9: Predicted curves for patients starting in the low, intermediate and high bilrubin states, from the 4 state and the 28 state AJ fit.

An analogy is to think of 3 teams at level A, B, C and an owner who aggressively moves players up and down based on some metrics, each thus has a rotating roster. Both the standard KM and the multistate attempt to estimate the long term outcome for a player who begins the season at level k. The EKM estimates the long term performance of *team* A, B, or C, which is something quite different.

Multiple covariates

The full PBC risk score model has 5 covariates, 4 of them with sequential measurement. The next most important variable after log(bilirubin) is log(albumin), both in overall impact and in the change in $\beta/se(\beta)$ between the time-fixed and time-dependent model, followed by edema. For categorization, we will break the albumin values into 3 groups of normal (≥ 3.5), $3--3.5$ and < 3; edema is already categorical. This creates 27 separate progression states, plus death.

```
[1] TRUE
```

The fits above verify that if we constrain the coefficients and the hazard, the MSH model fits are the same as an ordinary time-dependent Cox model fit. What about predictions? The next plot shows the AJ predictions for the 4 state and the 28 state models.

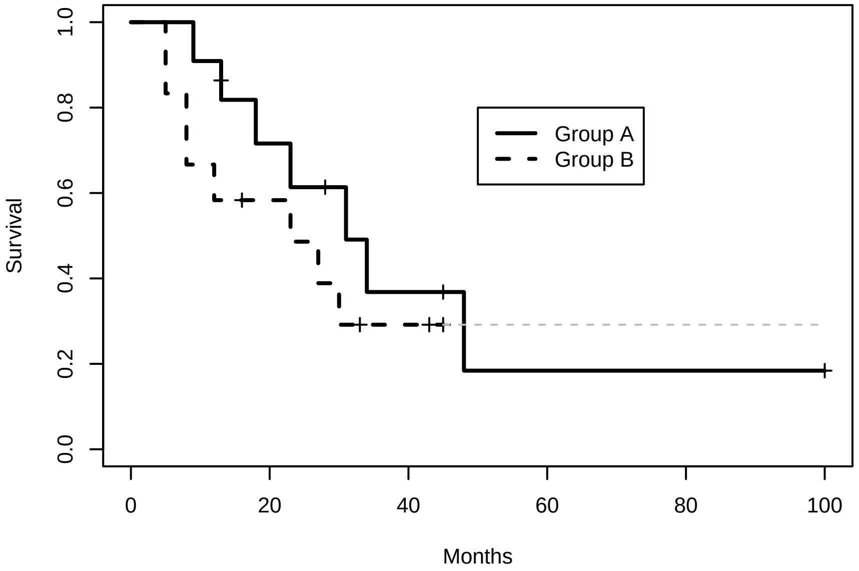

Figure 10: Survival curves from a small data set, with group B extended out to 100 months (light gray).

Empty states

Why is the 28 state AJ a below the 4 state AJ, for death, for those who start at high risk? An issue that will arise with any AJ fit where a state can become empty are artificial horizonal segments in the curve. This issue does not occur in a Kaplan-Meier curve because one normally stops plotting the data at that point. As an example consider figure 10, two survival curves from a small sample. The "predicted" survival for group B at 80 month requires extrapolation of the observed curve, and gives a value that is almost certainly biased upward: we would expect to see more deaths with further followup of the censored subjects in that group.

A multi-state Andersen-Gill estimate is subject to this same issue, but in a more subtle way. The software computes the estimates for all states jointly, until such time that all observations are censored. In the 28 state estimates of 9, for instance, if say the (bili 0–1, albumin ¿3.5, edema none) state becomes empty, p(death) will be underestimated from that point forward since it lacks contributions from that stratum, but the death curve may not have an obvious flat area due to contributions from the other 26 states. The curve for this (bili/alumin/edema) state itself will be horizontal while the state is empty, of course, but one is unlikely to plot all 27 of these. A strict procedure might be to terminate plotting for a given state whenever any of its inputs becomes empty due to censoring. In the case of the four state model, this will terminate the curves at year 14.2 instead of 14.3, so there is essentially no impact there, for the 28 state AJ the first such occurs just before 1.5 years (day 533). A crude estimate of the potential effect of these effectively missing periods of risk is to first create an overall death rate of a state r_j = (number of deaths from that state) / (total observed tim at risk in the state). Then add up, for each missing interval, $r_j p_j(t)\Delta t$, where the last is the width of the interval. For the dotted curves in Figure 9 this sum is surprisingly modest, less than .01 for each.

3.2 Multistate Hazard models

We can repeat the same exercise with predicted curves from a multistate hazard model. This allows for instance the addition of enrollment age as a second covariate. This approach has further flexibility, as it includes an entire family of models. One is to use separate hazards and coefficients for each transition, and is specified as

$$\lambda_{jk}(t;x) = \lambda_{jk0}(t)\exp(X\beta_{jk}) \tag{4}$$

where X includes age and perhaps other covariates, but does not include bilirubin, $jk0$ is a shorthand for the jk baseline hazard, and β is a matrix with one column for each jk transition. (This is a slight abuse of notation: if there were 5 states "jk" does not imply 25 baseline hazards and 25 columns for β; only those pairs with at least one observed transition appear.)

A fully additive model would be

$$\lambda_{jk}(t;x) = (\lambda_0(t)e^{\gamma_{jk}})\exp(X\beta) \tag{5}$$

Here all transitions share the same coefficients for X, and the baseline hazards have been assumed to be proportional. This would normally be, in our opinion, an unpalatably strong pair of assumptions for a multistate model, but gives a notion of the breadth available. An even stronger assumption would be $\gamma_{jk} \equiv 1$, that all the baseline hazards are identical; we have yet to encounter a data set where that is rational.

Note that γ is separate from β, the first is a coefficient that attaches to a transition, the second one attaches to a covariate. Nevertheless, by creating dummy variables for "current state", the coefficients γ and β can be jointly estimated using identical code to the MSH model. (As is often the case, the work is all in the data setup.)

Below we fit three variations.

1. Equation (4), with no covariates for transitions between bilirubin states, and enrollment age as a covariate for transitions to death. For the transitions between states 1, 2 and 3 the MSH baseline hazard will be completely non-parametric, which mimics the AJ hazard estimate. This represents a small expansion of the AJ by adding age as a covariate.

2. Constrain the above fit to have a common age coefficient for the 1:4, 2:4, and 3:4 transitions.

3. Further constrain to full additivity for the transitions to death per Equation (**??**).

	Age (β)			Bilirubin (γ)			Concordance
	1:4	2:4	3:4	1:4	2:4	3:4	
Model 1	2.2	2.5	1.4				
Model 2	1.6	1.6	1.6				
Model 3	1.6	1.6	1.6	1.0	2.9	27.1	

Table 3: Coefficients from 3 MSH model fits.

The coefficients and concordance values for the 3 fits are shown in Table 3. The shared coefficient for model 2 is a bit surprising until we note that the total number of transitions to death is 10, 22, 108 for bilirubin states 1–3, respectively; thus the 3:4 coefficient from model 1 carries the most weight in the combinded fit. The score test for model 1 versus 2 ($\chi^2 = 7.2$ on 2df, p = .02) argues against the simplification, as do the concordances. The coefficients for model 3 are identical to those from the simple time dependent model with age and bilirubin group as covariates, which is expected since these two fits share exactly the same form for the hazard of death. An overall score test between models 2 and 3 is not available, since they do not share the same baseline hazards. As shown in Figure 11, however, the predicted absolute risks hardly differ between the AJ and the MSH models adjusted for age. This is partly due to the fact that although age is highly significant in the model (1.6 fold hazard for each decade of increase, $p < .0001$) there is almost no correlation between enrollment age and initial bilirubin level. As well, proportional hazards fits this particular data very well. Remember that at the time of the study there was no effective treatment for PBC, and the study enrolled subjects across a wide range of ages and stage of disease.

Figure 11 shows predicted survival from the KM, AJ, and MSH models. For the MSH fits we have computed marginal estimates over the age distribution at enrollment. That is, 312 curves starting with normal bilirubin and enrollment age a_i, then averaged at each time point, repeat for moderate and high bilirubin at initial visit. There is very little difference, in this case, between the AJ and MSH curves, likely related to the excellent fit of a PH model to the PBC data. Remember that although the MSH hazards are proportional, $p(t)$ curves will not be so.

Figure 12 compares curves from the 4 state and 28 state MSH models. Unlike the AJ models in Figure 9, there is not a great difference in the prediction between 4 state and 28 state models. We (tentatively) ascribe this to the fact that the use of a shared baseline hazard for transitions to death precludes the issue of empty risk sets.

3.3 Marginal estimates.

There is a close conceptual connection between marginal estimates as discussed in Chapter **??** and the multistate models used above. Assume for a moment that we have a population of n subjects, without time-dependent covariates. Compute the n predicted survival curves $S_i(t)$, then the marginal estimate over

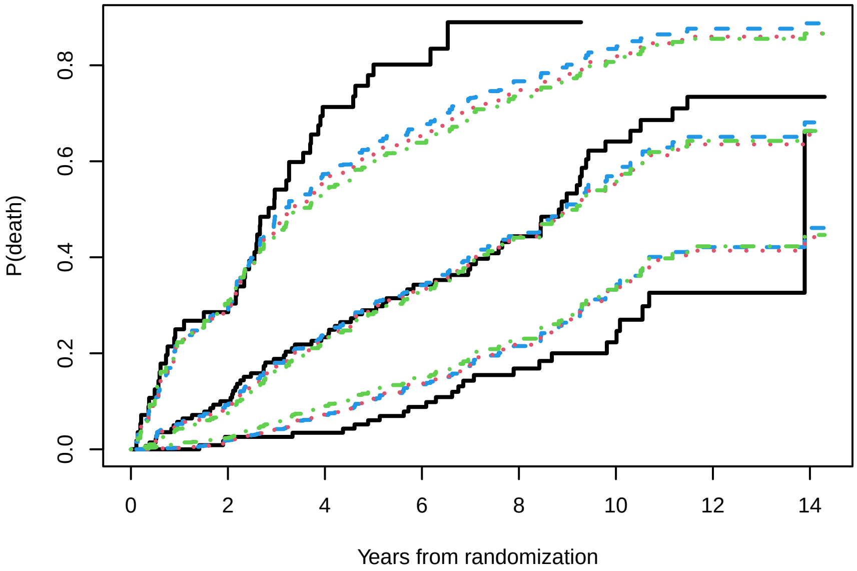

Figure 11: KM survival for the PBC dataset by initial bilirubin (solid black), along with predicted survival from the multistate AJ fit (dashed) and from MSH models 1 and 3 (dotted, dotdash).

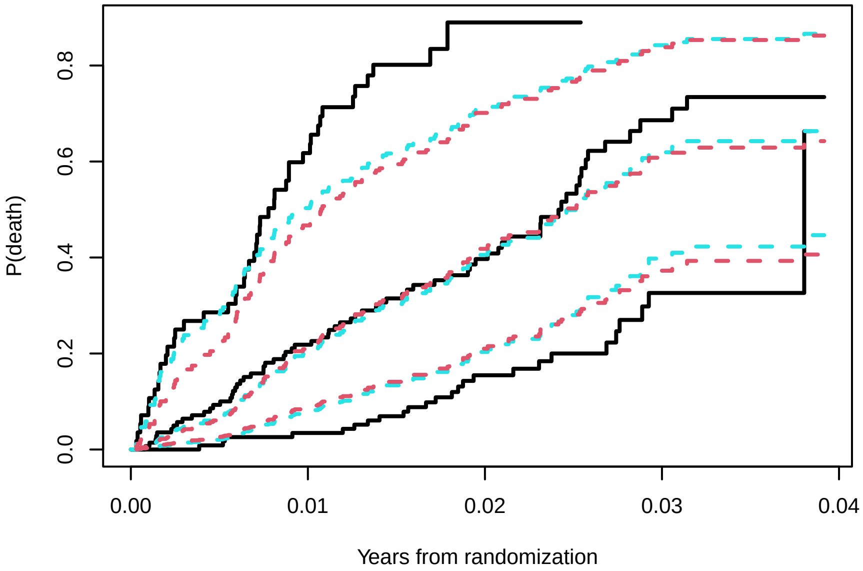

Figure 12: Predicted curves for patients starting in the low, intermediate and high bilrubin states, from the KM (solid), and the 4 state and 28 state MSH fits (dashed blue and red).

this population is

$$\begin{aligned}
\overline{F}(t) &= \sum_{i=1}^{n} F_i(t) && (6)\\
&= \sum_i \int_0^t f_i(u)du \\
&= \sum_i \int_0^t (1 - F_i(u)) \frac{f_i(u)}{1 - F_i(u)} du \\
&= \int_0^t \sum_i S_i(t) h_i(t) && (7)
\end{aligned}$$

where f is the density and h the hazard function. The last equation above is a weighted average of hazards at each time point. In a multistate model the inner sum becomes $p(t-)H(t)$ where H is the matrix of hazards.

From an ordinary Cox model fit using time-dependent covariates, one can create the predicted hazard $\lambda(t, x(t))$ for any pre-specified covariate path $x(t)$. This is, in fact, exactly what would be done for an external covariate path such as $x(t)$ = current age. This raises the idea of a time-dependent marginal estimate. That is, for each of the 116 subjects in the PBC study who enroll with bilirubin ≤ 1 (normal), use their individual observed bilirubin trajectories to create 116 such curves, and then average them to get an overall curve.

The problem with this is time dependent covariates and censoring. PBC subject 7, for instance is censored at 6.8 years. If we want estimates out to 12 years as in Figure 11, extending this particular subject's curve fronm 6.8 to 12 requires an extention of their time-dependent bilirubin values. By instead using a weighted sum of hazards as in Equation (7), the multistate approach allows use of hazards for those still under observation. Similar to the redistribute-to-the-right algorithm, subject 7 is represented in the future by the further transitions for all other subjects in the same state as subject 7 at year 6.8 (which happens to be bilirubin 1–4). The validity of the approach follows from the Markov assumption.

This also highlights one of the trade-offs with the multistate approach. Say that we desired more specific predictions, say for the cohort with enrollment bilirubin of 1.4–1.5, and to that end created 140 bilirubin states of [.2–.4), [.4, .6), . . . [27.8–28.0]. In the course of the computation, 10/172 (6%) of the observations will have no one else in their state at the time of censoring. The way that baseline hazards and/or the AJ are computed, such a subject is effectively imputed to have 0 chance of death from that point forward ($p(t)$ curves extend horizontally to the right), until and if that cell is repopulated by other transitions. The result is a biased estimate. (With bin widths of .1, 28% of the censors are orphaned in this way.)

Comment: I have thought about using continuous bilirubin. I'm not yet sure it is a great idea or a terrible one. The $p(t)$ function for the MSH has the same structure as that for the AJ, with an increment at time t of

$$p(t-)H(t)$$

where $H(t)$ is the transtion matrix, while p carries along, via $p(0)$, the information on where the curve started. The transition matrix is estimated using all subjects current at risk at time t — that is all we have. When computing H we could use continuous bilirubin, if we wished, to compute the individual λ_{jk} that form the elements.

- This needs a slightly different Markov assumption, that the distriution of covariates within the state is a surrogate for the covariate that left at censoring. Better/worse?
- I suspect it will help the edge effect. When someone transitions to the moderate bili state they take their "just barely over 1.0" value with them.
- This realization sets the stage for simple "margin" argument to survfit, i.e. it returns the marginal curve directly. That won't happen for a bit.

References

[1] E. R. Dickson, P. M. Grambsch, T. R Fleming, L. D. Fisher, and A. Langworthy. Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology*, 10:1–7, 1989.

[2] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, New York, 1980.

[3] Arvid Sjölander. A cautionary note on extended Kaplan-Meier curves for time-varying covariates. *Epidemiology*, 31:517–22, 2020.

[4] S. M. Snappin, Q. Jiang, and B. Iglewicz. Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. *Amstat News*, 59:301–7, 2005.

[5] T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model.* Springer-Verlag, New York, 2000.