Analysis of time-to-event for observational studies: Guidance to the use of intensity models, online supplement

STRATOS topic group 8

2020-08-21

1 Introduction

This is an online supplement to the STRATOS guidance paper for intensity models. It contains further background information on each of the three data sets, along with R code and discussion for all of the items in the checklist. A subset of this forms the results and figures for the examples section of the main paper. The files "intensity_supplement.pdf" (this file), "intensity_supplement.Rmd" "refer.bib" and "pad.rda" are included in the supplementary material for the journal and can also be found on the author's web site: http://publicifsv.sund.ku.dk/~pka/STRATOSTG8/.

No claim is made about whether R is the best language for this or other survival analyses. However, reproducible code in at least one package is an important component of the STRATOS document. The .Rmd file which generates this document can be found on the site, it reveals all the code and can be further explored using R or Rstudio. Execution of the .Rmd file will require the `survival` (version 3.1 or greater), `dynpred`, and `timereg` packages.

In some places the supplement will discuss alternative computations to those presented in the paper; these are cases where multiple variants can be used to accomplish the checklist goals; in each case our list will be far from exhaustive.

2 Peripheral Arterial Disease

2.1 Background

Peripheral arterial disease (PAD) is a common circulatory problem in which narrowed arteries reduce blood flow to peripheral limbs, often the legs. It is also likely to be a sign of a more widespread atherosclerosis, and subjects manifesting the disease carry an increased risk for all atherothrombotic events including myocardial infarction, stroke, and cardiac death. The `pad` data set contains the results of a Slovene study reported in Blinc et. al [Blinc et al., 2017, Blinc et al. [2011]]. Briefly, the study was conducted by 74 primary care physicians-researchers, each of whom was asked to recruit 10 subjects with PAD along with 10 age and sex matched controls without PAD; actual recruitment ranged from 1 to 31 pairs. All participating subjects consented to a yearly examination and interview for 5 years. Important endpoints are death, either due to cardiovascular disease (CVD death) or other causes, non-fatal CVD endpoints of infarction and stroke, and patient interventions attributed to the disease such as revascularization procedures.

The final study includes 742 PAD and 713 controls, with baseline data for each subject, measurements at each visit, and endpoints. This analysis will focus on cardiac endpoints and on overall mortality.

2.2 Endpoint and follow-up

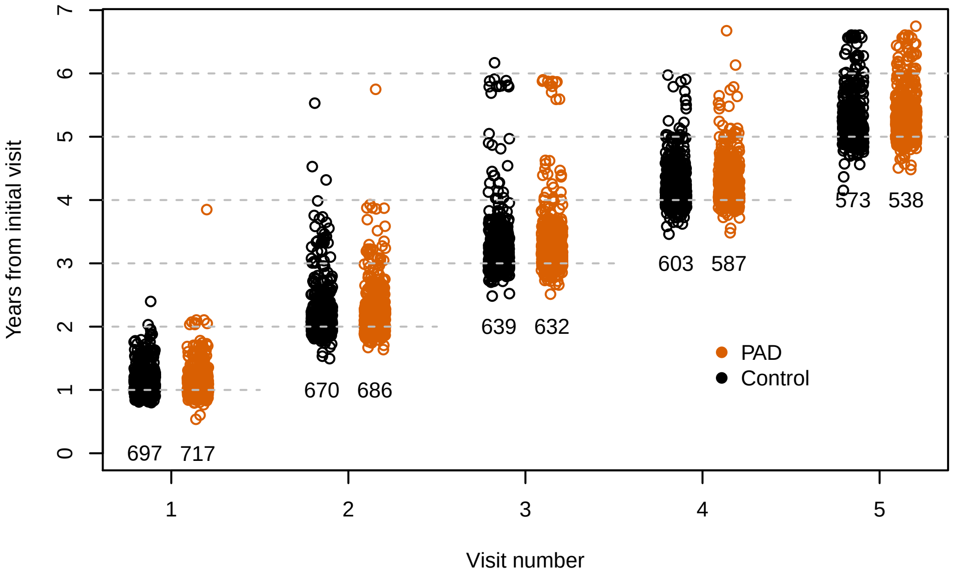


Figure 1: Time to each visit for the PAD study.

Before fitting covariate effects we look at the overall follow-up for subjects, and decide on a time scale and an endpoint. Per the study design subjects should have a visit at years 1–5. We see in Figure 1 that visit 1 dates are stretched out over a full year and by visit 5 over almost 2 years, a pattern that is unavoidable in observational studies of human subjects. Based on the counts shown in the figure there is no evidence of differential drop out between the PAD and control subjects. Figure 2 shows the censoring pattern for subjects in the two arms, computed using a reverse Kaplan-Meier. There is no evidence of differential follow-up. Given the visit slippage of the first figure, an ideal censoring pattern would be a final followup that is approximately uniformly distributed between 4.8 and 6.8 years, this is shown as a dotted line on the figure. Per the counts on Figure 1 the total *number* of visits for the subjects is not what was planned — about 25% do not have a fifth visit — but patients have adhered to the 5 year time commitment quite closely. As stated in the primary paper, a decision was made to truncate all patient data at 5 years for the purpose of analysis.

The starting R data has 3 dataframes:

- `pad1` has one observation per subject containing baseline information,
- `pad2` has one observation per subject visit containing measurements at that visit, and
- `pad3` has an observation for each observed endpoint.

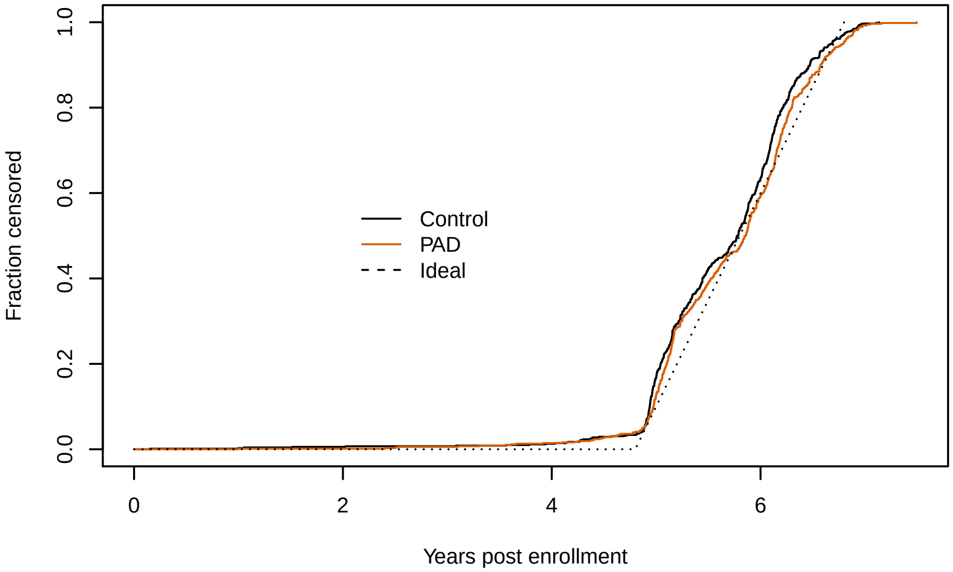


Figure 2: Observed and ideal follow-up in the PAD study

The initial survival curves can be created using a copy of the baseline data, with follow-up truncated at 5 years.

```
   censor    CV death nonCV death
     1296          68          91
```

Our first analysis will use time since enrollment as the primary time scale. For the PAD patients the time since diagnosis is a natural scale since it represents both the progression of disease and of treatments for the disease. Survival curves for the control subjects serve as a comparison outcome of similarly aged subjects without the disease, but do not have a natural stand-alone interpretation. Figure 3 contains the overall Kaplan-Meier for PAD and control subjects, male and female. Death rates are higher for males than for females, which is no surprise given a mean age at entry of 65 years, and is higher for PAD subjects than for the age matched controls. The right hand panel shows the curves on age scale.

Survival curves can equivalently display either the fraction who have died (curves start at 0) or the fraction who have not died (curves start at 1); the choice between the two is often based on tradition or habit. Pocock et. al. [Pocock et al., 2002] looked at the issue in a more scientific way, and recommend using curves that start at zero as the most reliably informative, particularly when the the overall death rate is < 30%; we will follow that advice.

2.3 Covariates and Cox model

The analysis will focus on the covariates of PAD, age, sex, and serum values for high density and low density lipoprotein (HDL, LDL). Figure 4 shows density estimates for each of the three continuous covariates at baseline, by PAD group. By study design there should be no difference in the age distribution. HDL and LDL are slightly lower for the PAD subjects.

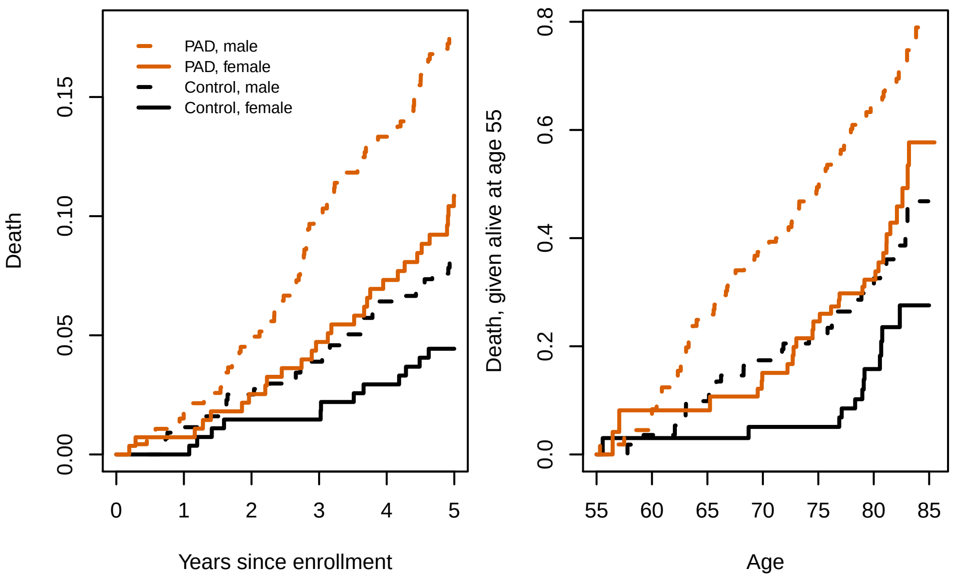


Figure 3: Kaplan-Meier curves of overall survival.

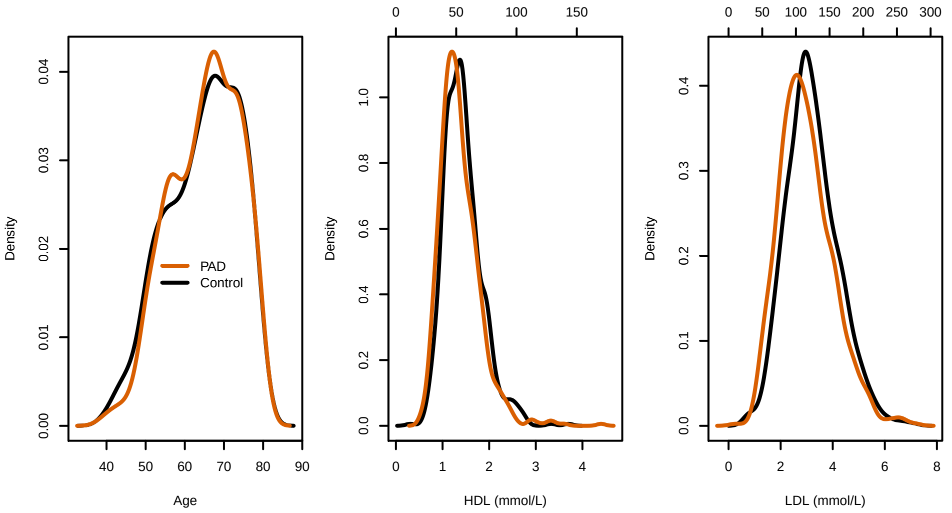


Figure 4: Distributions of three continuous covariates, by group. Blood value scales are in mmol/L (lower) and mg/dl (above).

The Cox model fits will use age, sex, baseline values of high density and low density lipoproteins (`hdl0, ldl0`), time-dependent lipid values (`hdl, ldl`), and a time-dependent value of "years since enrollment" (`etime`). Time since enrollment only plays a role in age scale models: in a time-since-entry model every one in a given risk set will have exactly the same value of the covariate, and hence the estimated coefficient for the variable would be 0.

For time-dependent analysis we use a counting process form. Mathematically each subject is represented by 3 processes: $Y_i(t) = 1$ if the subject is currently at risk, $X_i(t)$ is the time dependent covariate process, and $N(t)$ the cumulative number of events. In the R code this translates into multiple observations for each subject of the form

```
 id, time1, time2, event, x1, x2, ...
```

where `(time1, time2)` describe an interval over which the subject is at risk, `event` corresponds to $dN(t)$ and will be 1 of the subject had an event at `time2` and 0 otherwise, `x1, x2, ...` are the value of the covariates over the interval, and `id` is a subject identifier. Subject 9, for instance, is a male with blood values measured on days 0, 583, 945, and 1309, who died on day 1349. Their data has the form

```
id tstart tstop death  sex  ldl  hdl
 9      0   583     0 male 2.66 1.08
 9    583   945     0 male 3.20 1.53
 9    945  1309     0 male 2.81 1.32
 9   1309  1349     1 male 2.51 1.30
```

The lipid values and time intervals encode the most recent measurement of the value. This representation of counting process data is available in R, SAS, Stata and many other packages that deal with survival data. (It first appeared in the S package in 1987.) Creation and checking of the data often represents a major share of the work in an analysis. There is often concern that because the data set has multiple rows for the same subject, a valid variance estimate for the model will need to account for correlation.
This is not the case, however.
Internally, there is only one copy of each subject; the multiple rows are merely a way to communicate to the program *which* covariate value is applicable at any given time point.

When there are competing risks, such as CVD and non-CVD death, the 0/1 event event variable is replaced by a factor with levels of "none", "CVD death", "non-CVD death", which shows both when an event occurred and it's type (this feature is available in version 3.1 or later of the R survival package). Below we create separate time-dependent data sets for analysis on the time-from-enrollment scale (`pdata2`), the age scale (`pdata3`), and for the Poisson model (`pdata4`). (A normal analysis would choose only one of these, of course.) The two data sets have separate event indicators for death, the competing risk of CVD/non-CVD death, and for any CVD event/non-CVD death. Since the technical details of how to create the data sets are not a primary focus of this document the code to do so is not shown by default in the printed form, but as mentioned in the introduction it is present and viewable in the `online.Rmd` source file.

2.3.1 Overall mortality

Results of the fits are either shown below or are gathered together in Table 1. The first pair of fits shows that the estimated overall effects of PAD and sex hardly differ between analyses

Table 1: Coefficients from the Cox and Poisson models

	Overall		PAD		Control		p
	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)	PAD vs C
Time since enrollment scale							
PAD	2.40	(1.71-3.37)					
male sex	2.00	(1.40-2.86)	2.01	(1.31-3.08)	1.97	(1.02-3.79)	0.96
10 yr age	1.93	(1.57-2.37)	1.91	(1.49-2.45)	1.98	(1.36-2.89)	0.88
Age scale							
PAD	2.40	(1.70-3.37)					
male sex	2.02	(1.42-2.90)	2.01	(1.31-3.08)	2.01	(1.04-3.88)	1.00
yrs since entry	1.18	(1.05-1.33)	1.20	(1.05-1.38)	1.12	(0.91-1.39)	0.61
Poisson							
PAD	2.38	(1.70-3.35)					
male sex	2.01	(1.41-2.88)	2.03	(1.33-3.11)	1.97	(1.02-3.81)	0.95

done on the enrollment or age scales. A second pair fits looks at the coefficient effects within the PAD and control groups. (Inclusion of the covariate by strata interactions gives results that are equivalent to separate fits for the the PAD and control groups; bundling them as a single model allows a formal test of whether the effects differ between the two strata.) None of sex, enrollment age, or time since enrollment show meaningful differences between the PAD and control groups. That is, on the log-hazard scale, PAD appears to act as an additive risk.

An approach that allows for dual time scales is to use a Poisson regression model. To fit the model first categorize each subject's follow-up time into a set of discrete bins based on current age and and enrollment time; We will use single single years for time since enrollment and 5 year age groups. A Poisson model is then fit with the 0/1 death variable as the response, a separate intercept for each time/age strata, and the length of the time interval as an offset. This leads to $12 * 5 = 60$ separate intercepts in the model; to conserve space these are omitted from the printout. (Because some cells have 0 subjects < 60 terms appear in the final fit).

2.3.2 Cardiovascular death

Since PAD is expected to exert its primary effect via cardiovascular disease (CVD) endpoints, a more interesting analysis is to focus on CVD rather than overall mortality. Since in this case non-CVD death will be a competing risk this needs to be accounted for in our estimate of the probability in state; in the R survival package this is accomplished by using a multi-level categorical variable as the "status", rather than 0/1. The `survfit` function then computes Aalen-Johansen (AJ) estimates of a multi state model. The resulting curves are identical to "cumulative incidence" curves in a competing risks case (the CI estimator is a special case of the AJ estimate).

Figure 5 shows the Aalen-Johansen estimates of probability in state for CVD and non-CVD death in the left panel (which could also be labeled as "absolute risk", "competing risk",

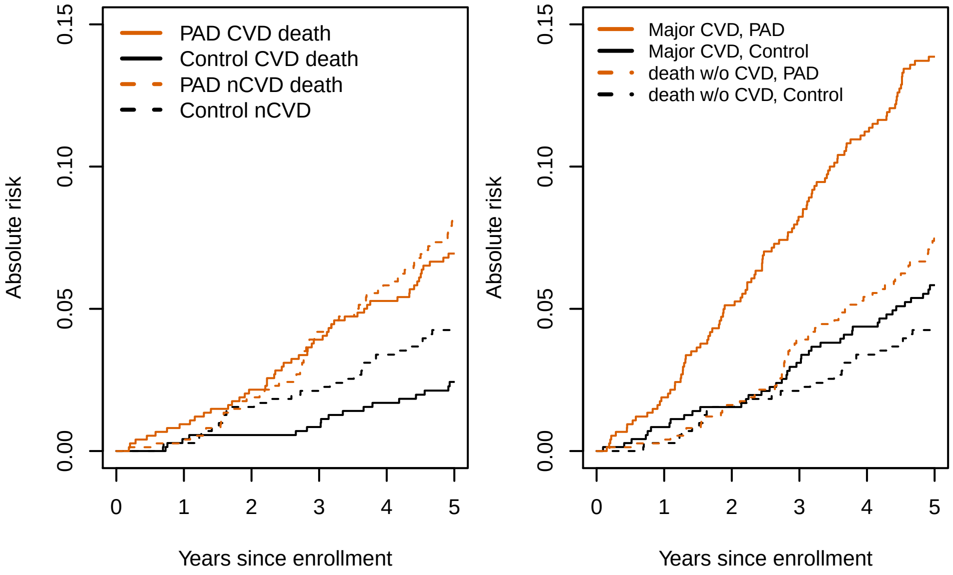


Figure 5: Aalen-Johansen curves for CVD vs. non-CVD death outcomes (left) and for any CVD event vs. death without CVD (right).

or "cumulative incidence") and the AJ curves for first major CVD event (infarction, stroke, CVD death) versus non-CVD death in the right panel. For controls, CVD deaths are about 1/2 of the non-CVD, approximately 2% versus 4% at 5 years; both outcomes are near 8% for the PAD subjects. Nearly 14% of the PAD subjects experience a major CVD event versus 6% for controls.

The absolute risk curves require that all transitions in the model be considered jointly: incorrect curves will result from a simple Kaplan-Meier of time to CVD, censoring at non-CVD death, or a KM of time to non-CVD death, censoring at CVD. Interestingly, this is not true for the cumulative hazard estimate: the naive estimate for one outcome, censoring the others, will be identical to the joint estimate. The Cox model is a model for the hazards and has the same identity: the first `coxph` call below, which models the entire process, will have the same results for the CVD death portion as the second `coxph` call, which ignores non-CVD deaths by censoring them; this is verified by the `all.equal` test. However, if one desires to predict absolute risk curves from the fitted Cox model then the first form is essential.

```
transitions table
             to
from           CV death nonCV death (censored)
  (s0)               67          91       1275
  CV death            0           0          0
  nonCV death         0           0          0
```

```
cat("Competing risks death, using baseline lipids\n")
print(pcox1)
cat("Competing risks death, using time-dependent lipids\n")
```

Table 2: Hazard ratios and 95% confidence intervals for the competing risks Cox models.

	Baseline lipids		Time-dependent lipids	
	HR	(95% CI)	HR	(95 CI)
CVD death				
PAD	2.87	1.01–8.14	2.40	0.84–6.84
male sex	1.67	0.59–4.70	1.36	0.49–3.83
age (10)	1.93	0.87–4.26	2.01	0.90–4.45
HDL (mmol/l)	0.74	0.24–2.28	0.21	0.06–0.75
LDL (mmol/l)	0.92	0.46–1.85	0.76	0.36–1.61
CVD event				
PAD	2.42	1.04–5.66	2.27	0.97–5.31
male sex	1.56	0.82–2.96	1.54	0.82–2.92
age (10)	2.09	0.86–5.06	1.90	0.79–4.57
HDL (mmol/l)	0.56	0.21–1.49	0.48	0.18–1.29
LDL (mmol/l)	1.09	0.62–1.93	0.88	0.48–1.62

```
print(pcox2)
```

There were 67 CVD and 91 non-CVD deaths in 5 years of follow-up, which are the transitions from the initial state (s0) to the CVD death and non-CVD death states, respectively, in the transitions table. No one transitioned from death to another state, which would indicate an error in the data set. Table 2 displays the coefficients for the fits with baseline and time-dependent lipids, for both of the competing outcomes. The joint Cox model shows that after adjustment for age, sex, and baseline lipids, PAD is associated with a 2.9 fold increase in the rate of CVD deaths and a 1.9 fold increase in non-CVD death. Age and sex are important predictors as well, as expected, with male sex having a somewhat larger effect on the non-CVD endpoint. The effect of baseline lipid values was not significant, however, time-dependent HDL shows an effect on the CVD death endpoint, somewhat reducing the estimated effect of PAD on CVD death.

2.3.3 Any CVD event

An alternative analysis is to use all major CVD events (stroke, infarction, CVD death) as the first endpoint and “death without a prior CVD event" as the competing endpoint. This increases the number of CVD events to 142, while the number of competing events is 85. At this point we are not interested in modeling the further transition from a non-fatal CVD event to death (15 events), and so exclude those with a prior CVD event from the analysis.

```
           to
from        CVD event death (censored)
  (s0)            143    85       1227
  CVD event         0    15         69
  death             0     0          0
```

The analysis on age scale and the analysis on time-since-enrollment scale give very similar estimated effects. The large impact of time since enrollment in the age scale models may be a surprise, but is actually a common finding for studies in non-acute settings that require patient consent. Essentially, patients who are in extremis do not volunteer. In the first six months after enrollment, it not uncommon to have a death rate that is less than 1/2 of the overall population death rate. The risk set at age 71, say, might have subjects who have been in the study for only a few months (recruited at age 71) and others on study for 4 years. In addition, the PAD study explicitly excluded subjects with a malignancy within the last 5 years.

2.3.4 Evaluating linearity

Below we reprise the CVD portion of the fit, and investigate linearity for the serum lipids, whose skewness may raise concerns. We can explore nonlinearity in several ways, in R fitting a spline is a simple one.

The likelihood ratio test provided by the `anova` command shows that the gain in goodness of fit from an HDL and/or LDL transformation is modest. The plot implies there is less "high HDL" benefit after reaching a value of 1.5, which agrees with common guidelines; together this suggests that a modified covariate `hdl2 = min(HDL, 1.5)` might be the better predictor. A final check for proportional hazards did not reveal any concerns.

```
Analysis of Deviance Table
 Cox model: response is  Surv(tstart, tstop, cvstat == "CVD event")
 Model 1: ~ pad + age10 + sex + hdl + ldl
 Model 2: ~ pad + age10 + sex + ns(hdl, 3) + ns(ldl, 3)
   loglik  Chisq Df P(>|Chi|)
1 -985.67
2 -980.93 9.4697  4   0.05037
            chisq df     p
pad         0.413  1 0.520
age10       3.249  1 0.071
sex         2.361  1 0.124
ns(hdl, 3)  6.453  3 0.092
ns(ldl, 3)  2.343  3 0.504
GLOBAL     12.787  9 0.172
```

2.3.5 Predicted probability-in-state (survival) curves

Predicted $P(state)$ curves can be created from the multi-state fits, as well. As with any Cox model prediction, the first task is to choose a set of covariate values for which prediction is desired. We create the predictions for four hypothetical male subjects with ages of 58 and 72 (quartiles of age), and PAD vs. control. For HDL and LDL use values of 1.3 and 3, which are near the medians. (These curves are for the model of CVD death vs other death, time-fixed hdl and ldl.)

```
      pad age10  sex hdl0 ldl0
1     PAD   5.8 male  1.3    3
2 Control   5.8 male  1.3    3
3     PAD   7.2 male  1.3    3
4 Control   7.2 male  1.3    3
strata   data states
```

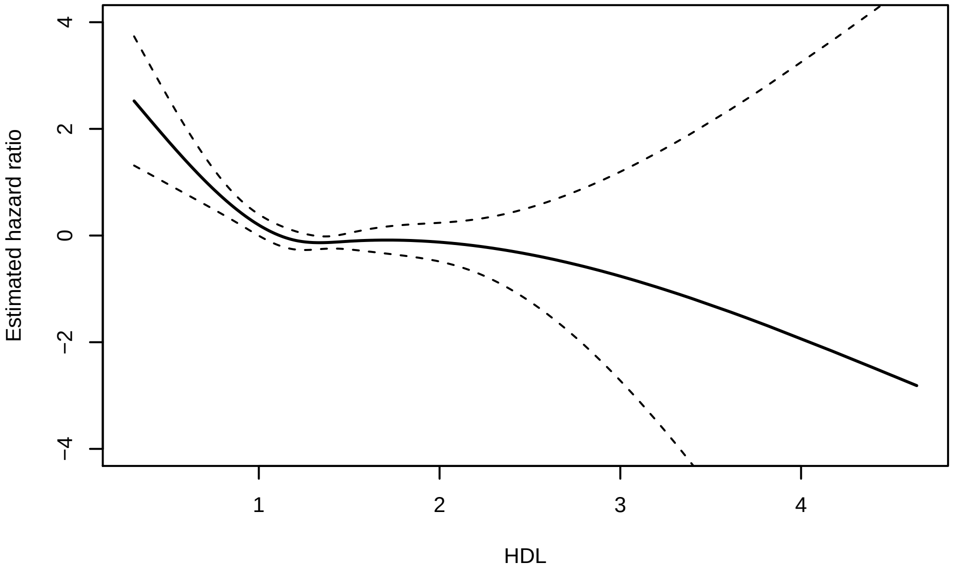


Figure 6: Estimated risk as a function of HDL level.

```
     1       4      3
[1] "(s0)"        "CV death"    "nonCV death"
```

The resulting set of absolute risk curves can be viewed as an array of estimates by stratum, target covariates, and state. The `pcox1` fit had no strata so the first dimension is 1. The resulting curves are shown in Figure 7. Because a multi-state model can generate a very large number of curves it is often helpful to break them into multiple panels.

3 NAFLD Study

3.1 Background

Non-alcoholic fatty liver disease (NAFLD) is defined by three criteria: presence of greater than 5% fat in the liver (steatosis), absence of other indications for the steatosis such as excessive alcohol consumption or certain medications, and absence of other liver disease [Puri and J., 2012]. NAFLD is currently responsible for almost 1/3 of liver transplants and it's impact is growing, it is expected to be a major driver of hepatology practice in the coming decade [Tapper and Loomba, 2018]. The `nafld` data set includes all patients with a NAFLD diagnosis in Olmsted County, Minnesota between 1997 to 2014 along with up to four age and sex matched controls for each case [Allen et al., 2018]. (Note that some changes to the public data have been made to protect patient confidentiality; analysis results here will not exactly match the original paper).

The goal of the study is to understand the progression of liver disease over time, whether NAFLD subjects are at increased risk for death or other endpoints as compared to the general population, and if so the amount of that increase. Because of the connection between NAFLD and obesity, sorting out actual causality is challenging. A further challenge is that

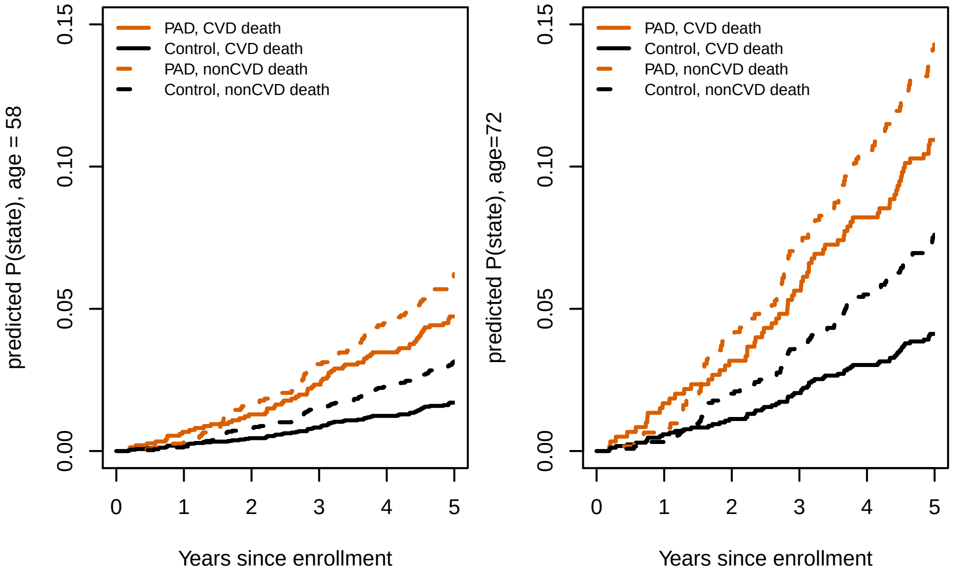


Figure 7: Predicted absolute risk curves from the multistate Cox model: dying of CVD (solid) and from other causes (dashed) by PAD status, for a male enrolled at age 58 (left panel) or 72 (right). The values of baseline HDL and LDL are 1.3 and 3, respectively.

only a minority of subjects are tested for NAFLD since this requires an abdominal scan. Thus we can only address the progression of *detected* NAFLD.

3.2 Endpoint and follow-up

Because of the very wide age range in the NAFLD study, along with the fact that NAFLD may well exist for many years before detection, the natural scale for the NAFLD study is subject age. The resulting Cox model will compare the death rate of a subject with NAFLD to non-NAFLD subjects of the same age. Age scale is also more natural for the control subjects than "time since being selected as random match".

Entry Time Subjects enter the study at the age of NAFLD diagnosis or selection as a control, whichever comes first. Because NAFLD is often a disease of exclusion, a NAFLD diagnosis followed shortly by the diagnosis of another liver disease is considered a false positive. The analysis data set is restricted to "confirmed NAFLD", i.e., if someone were diagnosed on 2001-06-20, the index date for confirmed NAFLD would be 2002-06-20, assuming that another liver diagnosis, death, or incomplete follow-up did not intervene. The follow-up of the matched control subjects also commences on the "confirmed NAFLD" date. This is important. If the matched subjects' follow-up were started on 2001-06-20 then the control has the opportunity to die during that first year while the case does not, leading to immortal time bias.

Since NAFLD is not an acute condition, the primary endpoints of morbidity and death are highly related to age, and the cases and controls match with each other on that scale.
This approach also mimics an idealized (but impractical) study which included the entire population from birth forward, with time dependent NAFLD as the covariate; and the use

of an ordinary Cox model would be natural in that case.

Inclusion Criteria: When selecting the controls for any given NAFLD case at age a, it is very important only to use information that was available at age a for the controls. We cannot exclude subjects who have too short a follow-up (die or censored before age $a + 2$ say), later have diabetes, or, most particularly, those who will later become NAFLD patients.
Each of these is a variant of immortal time bias. In this data set, 331 of the subjects selected as controls were diagnosed with NAFLD at a later age. Care must be taken at the time of analysis to correctly deal with these patients. The preliminary checks and figures will treat each subject's value at study entry as fixed, the hazard models will treat NAFLD as a time-dependent covariate.

Endpoint: The primary focus of this analysis will be all cause mortality. The subjects in the study are administratively censored at the end of 2017, when the data set was created. A small number will be enter and leave the catchment area due to migration, but above the age of 50 migration rates into or out of Olmsted county are very low [St Sauver et al., 2012], about 1% a year. Dates are not included in the public data for confidentiality reasons, so we cannot generate a censoring plot and relate it to enrollment.

Data: Create the data set with time-dependent covariate values for the 4 major covariates. The variable `nafld0` is the subject's NAFLD status at enrollment.

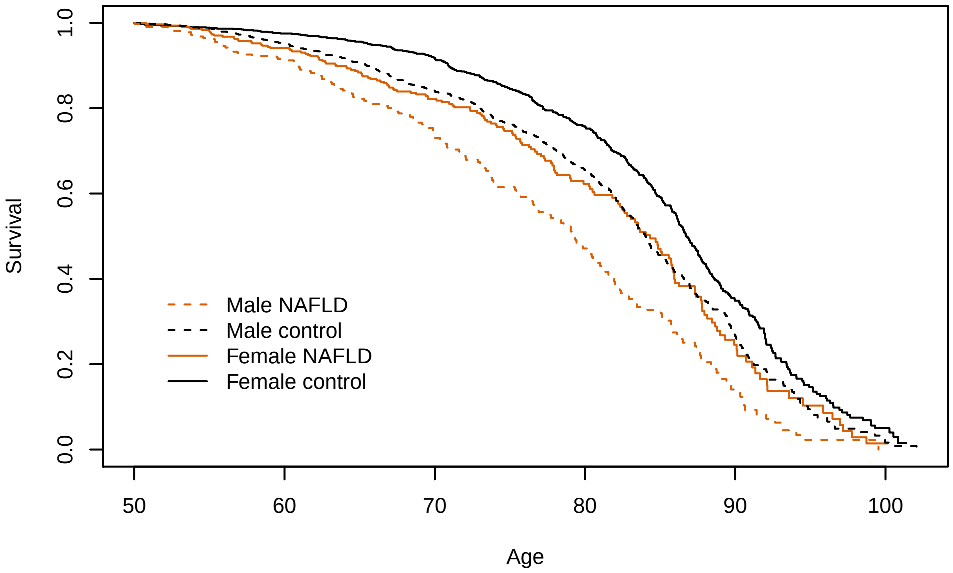


Figure 8: Survival curves from age 50 forward, comparing NAFLD to non-NAFLD at study entry, stratified by male/female.

Figure 8 shows the survival of the study stratified by sex and by NAFLD status at entry. Using NAFLD at entry is similar in spirit to using intent-to-treat in a clinical trial, in that it gives a reliable estimate but one that may underestimate the true clinical effect of a covariate. Data is plotted on the age scale. A primary message of the plot is that a NAFLD female has survival that is similar to a non-NAFLD male. Figure 9 shows the Nelson-Aalen estimates for the same data.

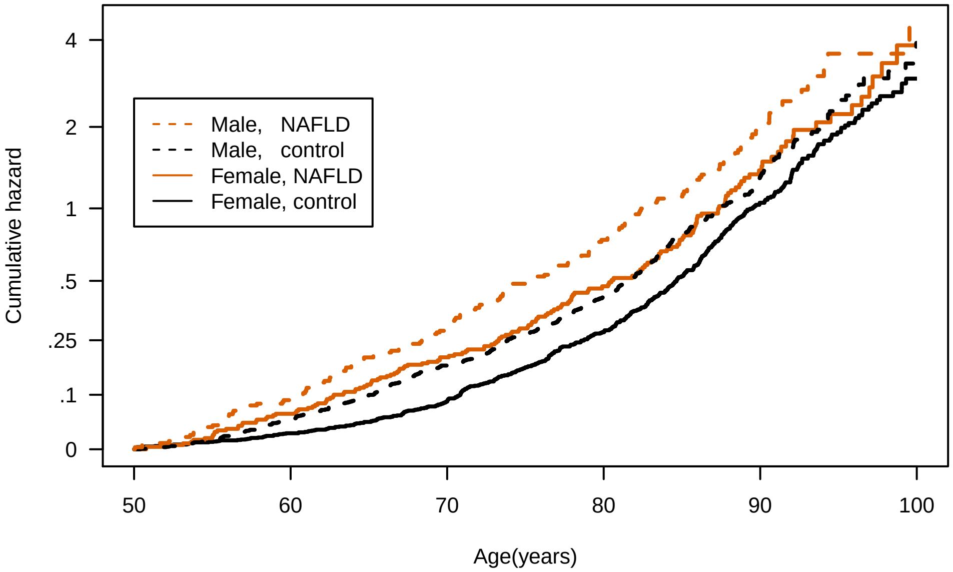


Figure 9: Nelson-Aalen estimates for the cumulative hazard from age 50, stratified by gender and NAFLD.

When using an age scale the left hand portion of a survival curve can often be highly variable due to small numbers, e.g. at the very first NAFLD age of 19 the study by definition only has 1+4 =5 subjects. This high early variability can then affect the entire curve. A strategy to combat this to to instead compute and plot $S(t|t > t_0)$, the future survival of those still alive at age t_0, with t_0 chosen so that the population at risk is “large enough”. The median age of NAFLD in the study is 52, we chose in the figure to plot $S(t|t > 50)$, the future outcome given survival to age 50, in part because we know a priori that population death rates before age 50 are low.

An alternate summary is to compute death rates by age group, sex, and time-dependent NAFLD status. The resulting table is closely related to landmark analysis, since changes in NAFLD status only apply forward in time. From ages 50 onward the death rates for a NAFLD female are close to those for a male control. Figure 10 displays the tabular data in graphical form.

```
      Female control Female NAFLD Male control Male NAFLD
<50              0.9          2.6          1.9       2.8
50-60            2.5          5.9          5.2       8.3
60-70            5.4         14.8         11.6      22.8
70-80           18.0         28.1         23.4      37.2
80+             82.7         92.3         95.3     124.4
```

4 Covariates and Cox model

For the Cox models the adjusting covariates of diabetes, hypertension and dyslipidemia are all time dependent. These adjusters are all related to the “metabolic syndrome”, which in

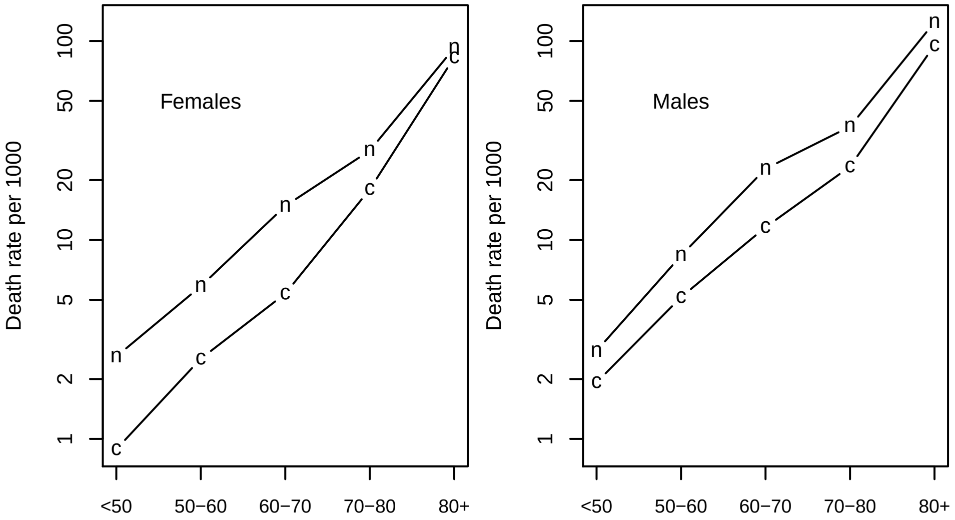


Figure 10: Death rates by age and sex, n = NAFLD, c = control.

turn is related to obesity. The table below shows the prevalence of each of them at entry to the study. Patients with NAFLD have higher rates of all the conditions.

```
             Control NAFLD
diabetes          10    28
hypertension      25    44
dyslipidemia      44    71
```

Now fit Cox models for death, stratified by male/female, with and without the adjusters, and then do separate fits for the males and females.

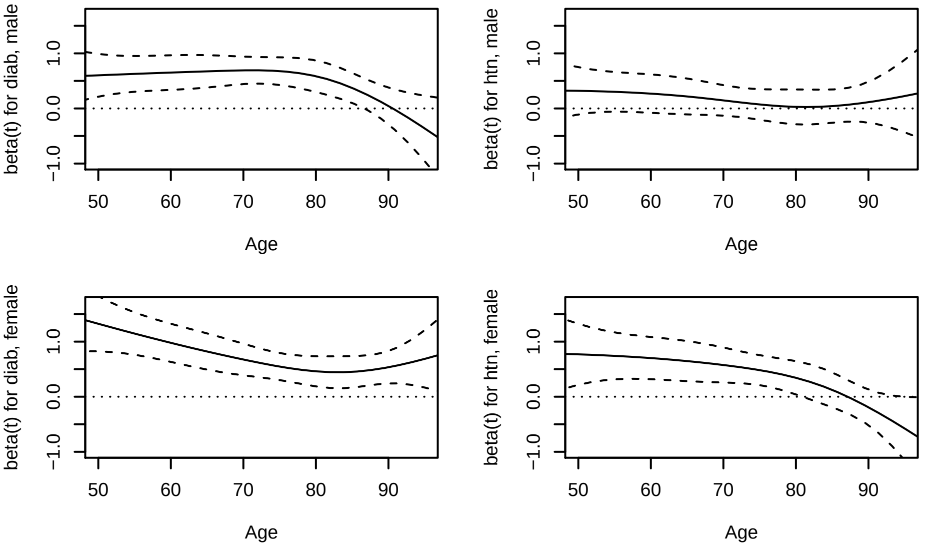
```
              All males females
NAFLD only   1.62  1.60    1.65
NAFLD        1.43  1.45    1.39
Diabetes     1.77  1.64    1.94
Hypertension 1.24  1.16    1.33
Dyslipidemia 0.68  0.72    0.65
```

The first line of the table estimates the effect of NAFLD on death, without adjustment for the covariates, while lines 2–5 show the multivariate model. First, the estimated effect of NAFLD is attenuated when adjusting for the three covariates. The higher prevalence of diabetes etc. explains a portion of the NAFLD effect. The overall NAFLD effect does not differ markedly for males and females.

```
       chisq df    p
nafld  0.739  1 0.39
GLOBAL 0.739  1 0.39
       chisq df      p
nafld   8.97  1 0.0027
GLOBAL  8.97  1 0.0027
```

The above shows tests for proportional hazards in the univariate model, separately for males and females, with the result being a strong indication against PH in the females but not in the males. These results are not surprising. In the preliminary plot of grouped rates shown in Figure 10 the curves for males are approximately parallel on the logarithmic scale, implying a constant ratio of hazards, whereas the female curves converge over time. More interesting is the result of PH tests on the multivariate model:

```
       chisq df      p
nafld  1.009  1 0.3150
diab   8.776  1 0.0031
htn    3.021  1 0.0822
dyslip 0.838  1 0.3601
GLOBAL 9.366  4 0.0526
       chisq df       p
nafld  11.20  1 0.00082
diab   20.77  1 5.2e-06
htn    22.07  1 2.6e-06
dyslip  1.18  1 0.27816
GLOBAL 37.16  4 1.7e-07
```



5 Advanced Ovarian data set

5.1 Background

The ovarian cancer data set contains follow-up on 358 subjects who were enrolled in two trials of chemotherapy for advanced ovarian cancer, conducted in between 19xx and 19xx by a multi-institution research network in the Netherlands. The eligibility criteria for enrollment included pathologic confirmation of advanced disease, age less than 70 years, lack of serious cardiac or renal disease, and favorable haematalogical status. Patients could not have a second tumor, brain metastases, or prior radiation or chemotherapy. The data is extensively analyzed in chapter 6 of [van Houwelingen and Putter, 2012], and further references can be found there as well.

Using covariates that were recorded at baseline it is desired to create a risk score that could be used to categorize the patients at that point. Patient follow-up in the data set continued for 6 years. Our analysis will focus on the following 3 covariates.

- FIGO: This is a staging system for ovarian cancer. Advanced ovarian cancer comprises the stages III (n=262) and IV (n=96). Stage IV patients have a very poor prognosis.
- The diameter of the residual tumor after surgery, categorized as micro, < 1 cm, 1–2 cm, 2–5 cm, and > 5 cm.
- Karnofsky index. A measure of the patient's functional status at the time of diagnosis. A score of 100% is an indication of no physical limitations.

5.2 Endpoint and follow-up

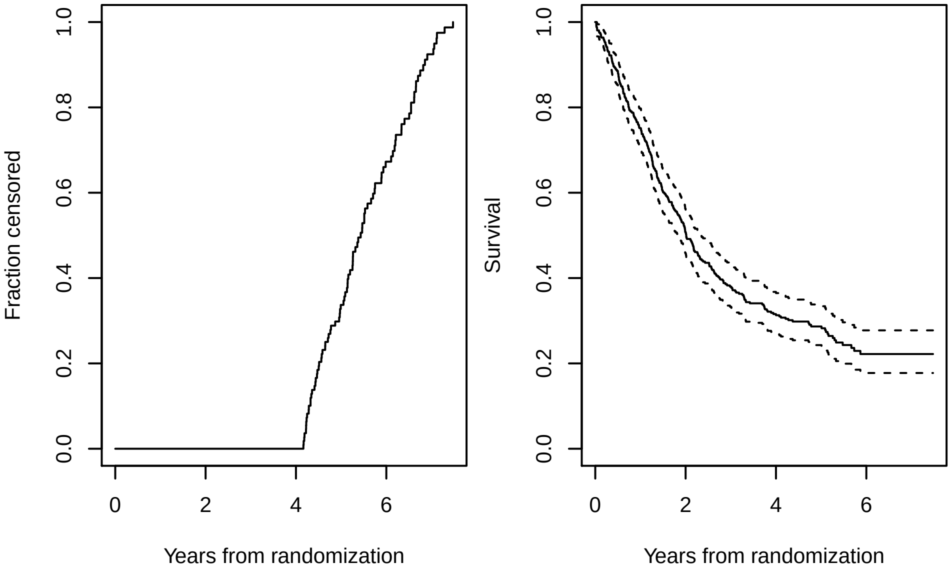


Figure 11: Censoring fraction and survival curves for the ovarian cancer study.

For a fatal condition such as advanced cancer, time from diagnosis forward is the most natural time scale, since it is the time frame of most direct interest to both the patient and the care provider. For a clinical trial such as this, the follow-up will be well-defined and regular.

The left panel of Figure 11 shows the censoring pattern for the study, which follows the expected "hockey stick" shape for a formal trial with 3 years of enrollment, 4 years of follow-up after enrollment of the final subject, and no subjects lost to follow-up. The graph shows no censoring before 4 years followed by an upward line corresponding to uniform accrual each year, with perhaps a slight flattening near the top right. Such flattening, when it occurs, normally corresponds to the earliest portion of the study when the accrual rate is still increasing.

The Kaplan-Meier gives the overall pattern of survival for this cohort. Inclusion in the study is based on each patient's current condition, and the endpoint of death is unambiguous.

Table 3: Data frequencies and univariate coefficients for the Ovarian Cancer data. (The coefficient is 0 for the reference category.)

Diameter				
micr.	<1cm	1-2cm	2-5cm	>5cm
29	67	49	68	145
0.00	0.42	0.93	1.02	1.23
FIGO				
III	IV			
262	96			
0.00	0.71			
Karnofsky				
6	7	8	9	10
20	46	47	108	137
1.17	0.81	0.31	0.07	0.00

For the ovarian study, the study timeline and endpoint are the simplest possible.

5.3 Covariates and Cox model

Table 3 shows the data distribution for three potential predictors, along with the coefficient from a univariate Cox model. For each of the categorical predictors the first level corresponds to the best expected outcome and is used as the reference level for the coding, while for Karnofsky score the highest value is used. Each of the variables' univariate effect is in the expected direction. Karnofsky score will be treated as continuous in the multivariate model – the categorical view in the table was for data examination.

An initial Cox model using all the covariates is below. The order of the variables in the model is based on their univariate significance.

```
ovcox1 <- coxph(Surv(tyears, d) ~ Diam + FIGO + Karn+ Broders + Ascites, ova)
anova(ovcox1)
Analysis of Deviance Table
 Cox model: response is Surv(tyears, d)
Terms added sequentially (first to last)

         loglik   Chisq Df Pr(>|Chi|)
NULL    -1414.3
Diam    -1395.5 37.6681  4  1.312e-07
FIGO    -1386.2 18.6508  1  1.570e-05
Karn    -1380.7 10.8876  1  0.0009681
Broders -1376.3  8.8128  4  0.0659532
Ascites -1374.7  3.1759  2  0.2043421
```

```
ovcox2 <- coxph(Surv(tyears, d) ~ Diam + FIGO + Karn, ova)
print(summary(ovcox2)$conf.int, digits=2)
```

```
          exp(coef) exp(-coef) lower .95 upper .95
Diam<1cm       1.38       0.73      0.73      2.57
Diam1-2cm      2.24       0.45      1.19      4.21
Diam2-5cm      2.38       0.42      1.29      4.39
Diam>5cm       2.53       0.39      1.40      4.57
FIGOIV         1.73       0.58      1.33      2.25
Karn           0.84       1.20      0.75      0.93
```

The most important covariates are the tumor diameter, FIGO stage, and Karnofsky index. Broders' grade adds only marginally to the first three and as seen above the univariate ordering was not clear, and ascites adds very little. For simplicity we will retain only the first three variables. The final model `ovcox2` has 3 predictors and 6 coefficients.

5.4 Checking proportional hazards

Per the checklist it is important to check proportional hazards. Because of a long research history on this topic there are a large number of ways to proceed with such a check, and some strong opinions about which of them is "best". We are more concerned that a check is performed rather than exactly how the check is done; below we will use the ovarian data set to illustrate a 3 by 2 matrix of common approaches. All of them are valid, and most packages have facilities for one or more. The first dimension of the choice is at what level to examine the proportionality assumption:

1. (highest): the risk score as a whole.
2. the effect of each term
3. the effect of each column of the X matrix.

For a continuous covariate levels 2 and 3 are the same. For a multilevel factor variable like Diameter level 3 asks if the *relative* hazard for a single levels of diameter changes over time, holding all the others fixed, while level 2 would be a multiple degree of freedom test. Level 1 can be assessed by using the overall risk score $\eta = X\hat{\beta}$ as a single covariate in a (new) Cox model. It may be useful as a first check but since an immediate follow-up question would be "which variable is the cause", we will not pursue it further.

The two tests we will use are

- Add a specially constructed time dependent covariate $x(t) = xg(t)$ to the model. This was suggested in Cox's original paper and so is available in nearly all packages. The rationale is that if $\beta(t) = a + bg(t)$ for some prespecified function g, then the coefficients a and b can be obtained by a fit that has both x and $z = g(t)x$ as terms in the model, and then testing if the constructed covariate z has a significant coefficient. An advantage of the test is its simplicity, the disadvantage is that $g(t)$ is arbitrary — should it be t, $\log(t)$, $\log(1 + t)$, …? A correct or near correct choice will have high power. The `cox.zph` function computes the score test version of this approach. It also can plot a semi-parametric estimate of $g(t)$ based on the method of Grambsch and Therneau Grambsch and Therneau [1994] that can help reveal outliers or other influential points. The plot may suggest an appropriate form for g.
- The cumulative sum of martingale based residuals method of Lin, Wei, and Ying Lin et al. [1993]. Since it is based on a maximum deviation (Kolmogorov-Smirnoff) the user does not have to make a choice of $g(t)$, making this the most robust choice. A

Table 4: Proportional hazards test statistics for the ovarian data.

	Score test		cumulative sums
	log(t)	rank(t)	K-S
Diameter (overall)	0.229	0.110	
Diam<1cm	0.337	0.184	0.152
Diam 1-2cm	0.474	0.630	0.286
Diam 2-5cm	0.103	0.053	0.128
Diam >5cm	0.690	0.663	0.946
FIGO	0.118	0.555	0.616
Karno	<0.001	0.001	0.028

disadvantage of the test is that it operates on each coefficient separately making the assessment of a categorical variable more complex.

The p-values for the tests are shown in Table 4, and the code to generate them all is below. We used both $g(t) = \log(t)$ and $\text{rank}(t)$ as test functions, the first is one of the more commonly used choices, and the second has is robust to outliers in time.

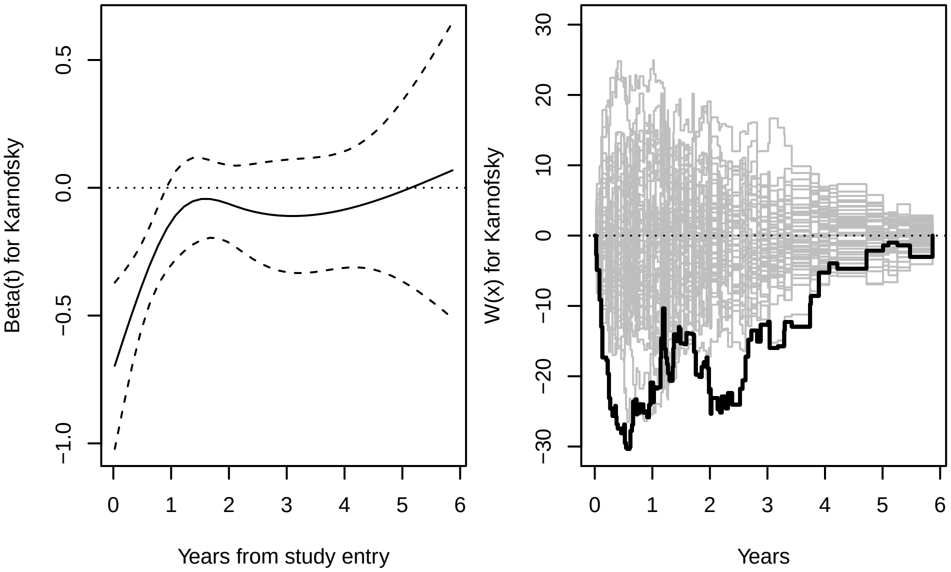


Figure 12: Proportional hazards plots for Karnofsky score based on the Grambsch and Therneau, and on the Lin, Wei and Ying tests.

For this data set, at least, the results show that all three statistical procedures lead to the same conclusion. Karnofsky score is the variable that is most implicated by the tests, and methods 2 and 3 both provide a plot to go along with the test. These are shown in Figure 12. In the GT plot proportional hazards corresponds to a horizontal line at the estimated Cox model coefficient. The plot for the GT method instead shows a rapid early drop in importance to near zero, implying that a 1 year old Karnofsky score is no longer predictive; something that is not surprising for advanced cancer. In the plot for the cumulative sums of residuals the gray lines are cumulative sums from a process under H_0 that the model assumptions hold and the black line is the process for the observed data. The Karnofsky line is far from the center of the cloud, again showing a failure of proportional hazards.

5.5 Dealing with lack of PH

One approach to deal with the lack of proportional hazards is to fit a set of *landmark* models. In this case we will break the data into three epochs of 0–2, 1–3, and 3-5 years, fitting a separate model in each. Each epoch then uses the covariate values available at the start of the epoch. (In this case there are no time-dependent covariates so all fits use the same values.)

```
          0-2 year 1-3 year  2-4
Diam<1cm      1.31     1.63 2.73
Diam1-2cm     2.92     2.89 2.17
Diam2-5cm     3.04     2.75 3.55
Diam>5cm      2.69     3.21 5.54
FIGOIV        1.76     1.70 1.64
Karn          0.77     0.89 1.07
```

We can predict the probability of survival for 1 more year for each of the three models, varying the target Karnofsky score and stage, while holding the diameter at < 1 cm. Figure 13 shows the result. In the first epoch there is a large difference between a Karnofsky 100 and Karnofsky 80 subject (black vs red) in predicted survival, but this difference shrinks in the 1–3 and 2–4 year epochs.
The effect of FIGO III vs IV (solid vs dashed), however, stays large across the time range.

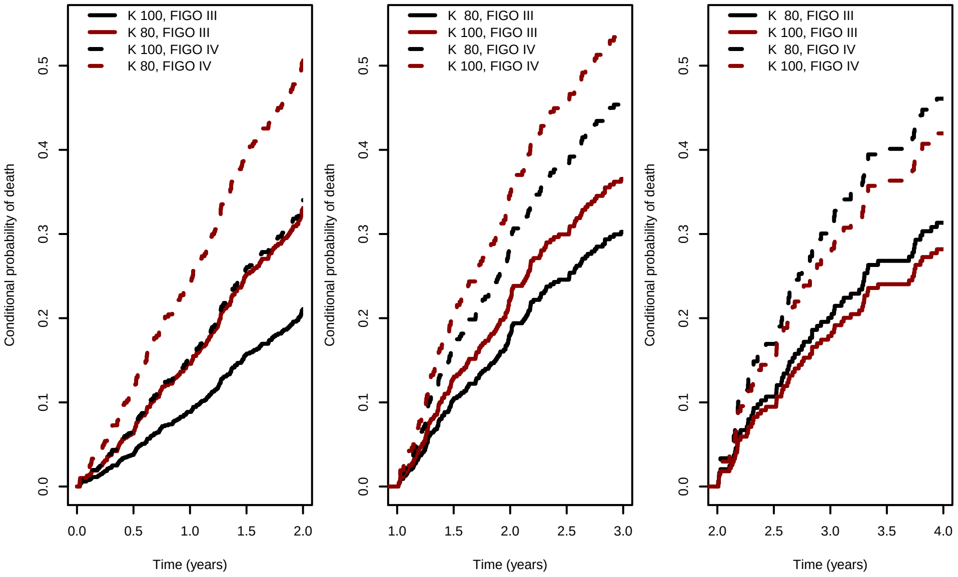


Figure 13: Conditional probability of dying in each epoch.

References

A.M. Allen, T.M. Therneau, J.J. Larson, A. Coward, V.K. Somers, and P.S. Kamath. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: A 20 year community study. *Hepatology*, 67:1726–36, 2018.

A Blinc, M Kozak, M Sabovič, M Božič, M Stegnar, P Poredoš, A Kravos, B Barbič-Žagar, M Pohar Perme, and J Stare. Prevention of ischemic events in patients with peripheral arterial disease – design, baseline characteristics and 2-year results, an observational study. *Int Angiology*, 30:555–66, 2011.

A Blinc, M Kozak, M Sabovič, M Božič, M Stegnar, P Poredoš, A Kravos, B Barbič-Žagar, J Stare, and M Pohar Perme. Survival and event-free survival of patients with peripheral arterial disease undergoing prevention of cardiovascular disease. *Int Angiology*, 35:216–27, 2017.

P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526, 1994.

D. Y. Lin, L. J. Wei, and Z. Ying. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80:557–572, 1993.

Stuart J. Pocock, Tim C. Clayton, and Douglas G. Altman. Survival plot of time-to-event outcomes in clinical trials: practice and pitfalls. *Lancet*, 359:1686–1689, 2002.

P. Puri and Sanyal A. J. Nonalcoholic fatty liver disease: Definitions, risk factors, and workup. *Clinical Liver Disease*, 1:99–103, 2012.

J. L. St Sauver, B.R. Grossardt, B. P. Yawn, L. J. Melton III, J. J. Pankratz, S. M. Brue, and W. A. Rocca. Data resource profile: The Rochester Epidemiology Project (REP) medical records-linkage system. *Int J Epi*, 41:1614–24, 2012.

E.B. Tapper and R. Loomba. Nonalcoholic fatty liver disease, metabolic syndrome, and the fight that will define clinical practice for a generation of hepatologists. *Hepatology*, 67: 1657–9, 2018.

H.C van Houwelingen and H. Putter. *Dynamic Prediction in Clinical Survival Analyis.* CRC Press, 2012.