Suvival workshop, part 4

Terry Therneau

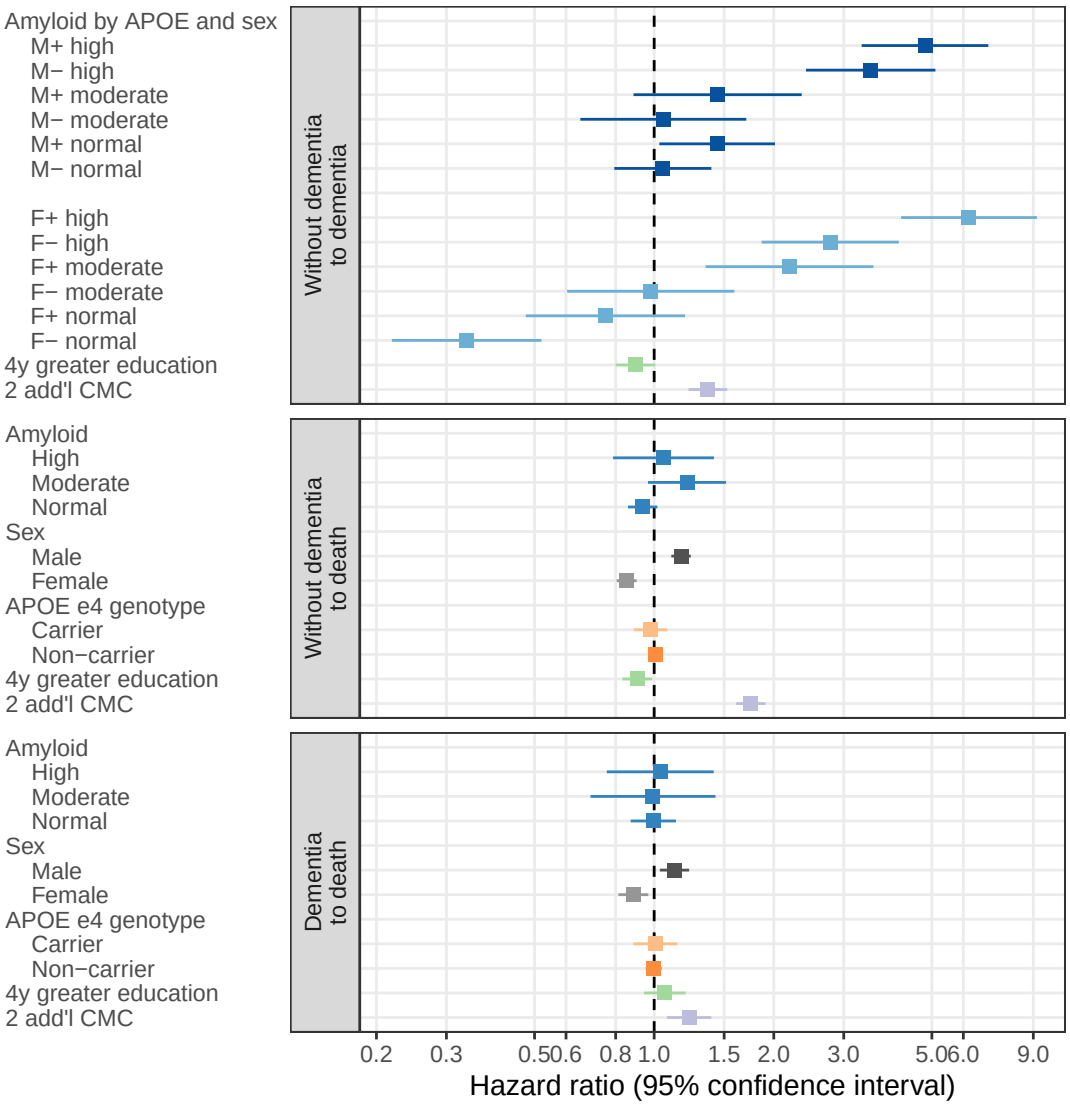
MCSA

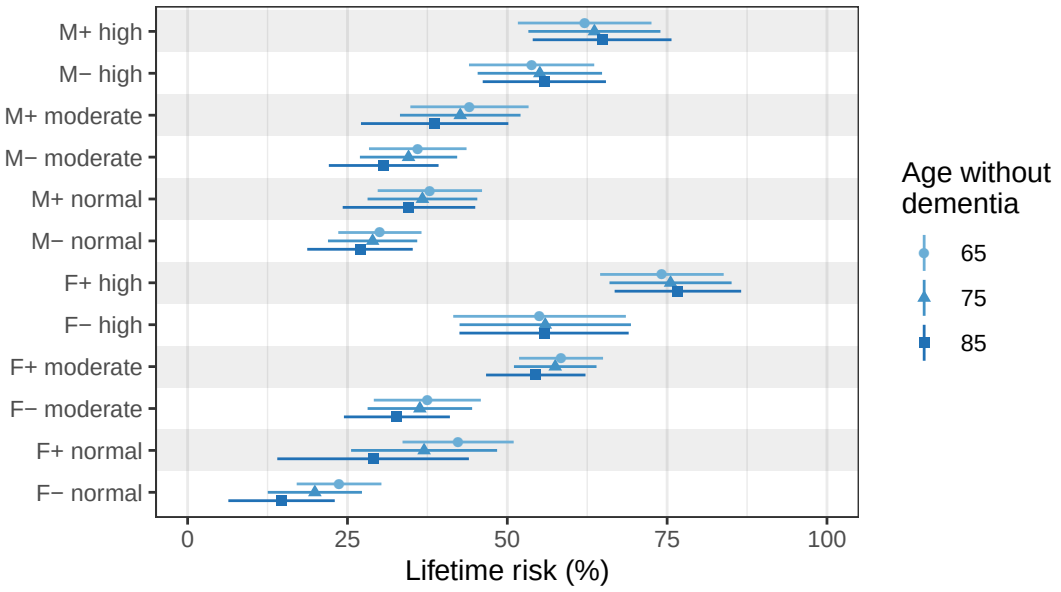
I have learned a tremendous amount about multi-state models from an analysis of the Mayo Clinic Study of Ageing (MCSA) data. One paper's analysis in particular (doi: 10.1093/braincomms/fcac017) has a directory with 14 different 'exporatory' knitr files.

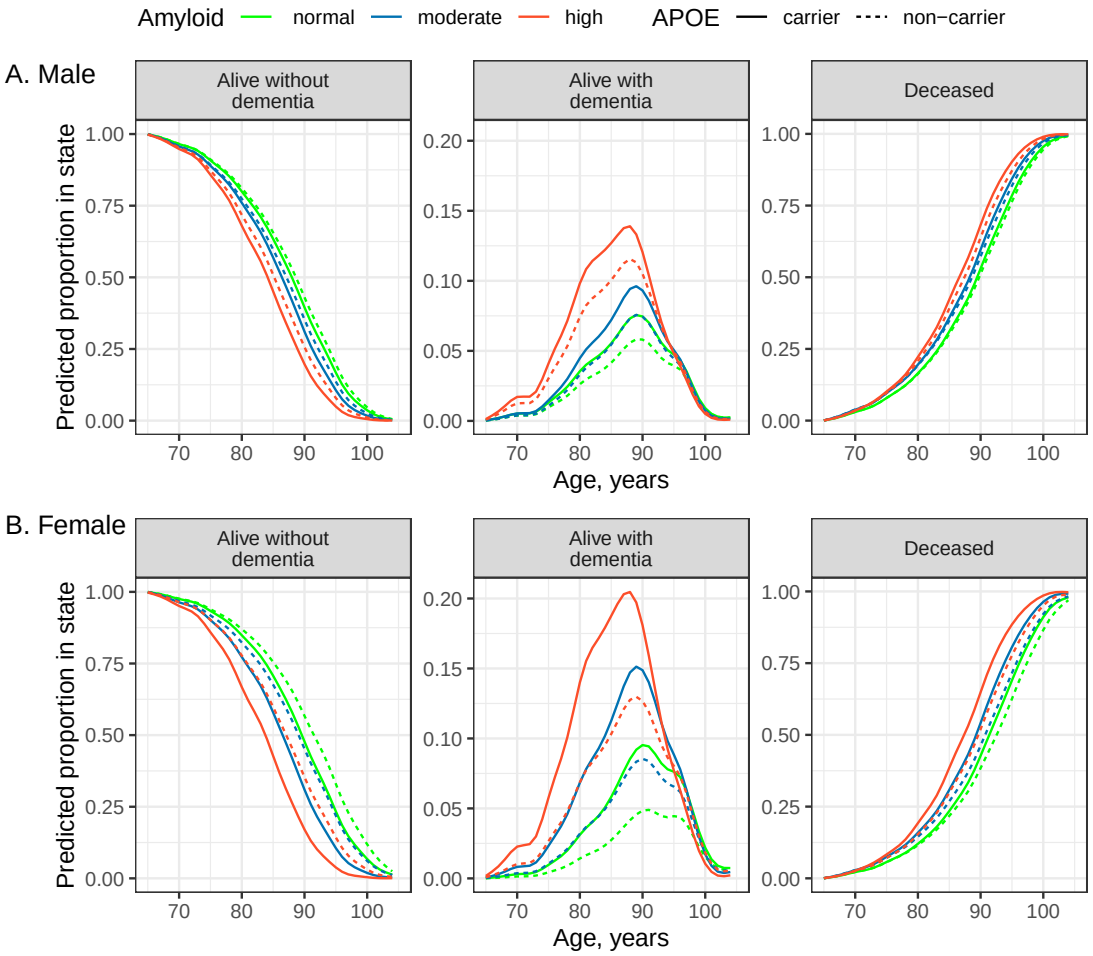
```
Call:
survcheck(formula = Surv(age1, age2, state) ~ 1, data = mcsa,
    id = ptnum, istate = cstate)

Unique identifiers       Observations        Transitions
              4944              63005               2400

Transitions table:
          to
from       dementia death (censored)
  ND            657  1203       3084
  dementia        0   540        111
  death           0     0          0

Number of subjects with 0, 1, ... transitions to each state:
          count
state         0    1   2
  dementia 4287  657   0
  death    3201 1743   0
  (any)    3084 1320 540
```

```
Call:
coxph(formula = list(Surv(age1, age2, state) ~ iclgp + apoepos +
    male + educ4 + icmc, 1:2 ~ iclgp:male + apoepos:male, 1:3 ~
    yearc), data = mcsa, id = ptnum, istate = cstate)


1:2                    coef exp(coef) se(coef) robust se    z     p
  iclgpmoderate       1.223     3.398    0.413     0.409  3.0 0.003
  iclgphigh           2.214     9.156    0.377     0.379  5.8 5e-09
  iclgpmissing        1.284     3.612    0.340     0.337  3.8 1e-04
  apoepos             0.846     2.331    0.114     0.116  7.3 3e-13
  male                1.318     3.736    0.372     0.365  3.6 3e-04
  educ4              -0.094     0.910    0.061     0.062 -1.5 0.126
  icmc                0.112     1.119    0.026     0.028  4.0 6e-05
  iclgpmoderate:male -1.287     0.276    0.506     0.505 -2.5 0.011
  iclgphigh:male     -1.122     0.326    0.441     0.446 -2.5 0.012
  iclgpmissing:male  -1.080     0.340    0.382     0.380 -2.8 0.005
  apoepos:male       -0.512     0.599    0.165     0.167 -3.1 0.002
```

```
1:3               coef exp(coef) se(coef) robust se    z      p
  iclgpmoderate  0.298     1.347    0.155     0.147  2.0  0.044
  iclgphigh      0.189     1.208    0.183     0.173  1.1  0.275
  iclgpmissing   0.542     1.719    0.102     0.098  5.5  4e-08
  apoepos       -0.029     0.972    0.069     0.067 -0.4  0.668
  male           0.356     1.428    0.060     0.059  6.0  2e-09
  educ4         -0.129     0.879    0.045     0.046 -2.8  0.005
  icmc           0.209     1.232    0.020     0.021 10.0 <2e-16
  yearc1        -1.059     0.347    0.157     0.158 -6.7  2e-11
  yearc2-5      -0.351     0.704    0.068     0.067 -5.2  2e-07


2:3               coef exp(coef) se(coef) robust se    z     p
  iclgpmoderate -0.039     0.962    0.254     0.225 -0.2 0.864
  iclgphigh     -0.043     0.958    0.220     0.212 -0.2 0.841
  iclgpmissing   0.073     1.075    0.179     0.173  0.4 0.675
  apoepos        0.051     1.053    0.093     0.084  0.6 0.539
  male           0.264     1.302    0.092     0.084  3.1 0.002
  educ4          0.083     1.087    0.064     0.060  1.4 0.168
  icmc           0.039     1.040    0.028     0.027  1.4 0.157

 States:  1= ND, 2= dementia, 3= death

Likelihood ratio test=444  on 27 df, p=<2e-16
n= 63005, number of events= 2400
```


Amyloid by APOE and sex
M+ high
M− high
M+ moderate
M− moderate
M+ normal
M− normal
F+ high
F− high
F+ moderate
F− moderate
F+ normal
F− normal
4y greater education
2 add'l CMC
Without dementia to dementia
Amyloid
High
Moderate
Normal
Sex
Male
Female
APOE e4 genotype
Carrier
Non−carrier
4y greater education
2 add'l CMC
Without dementia to death
Amyloid
High
Moderate
Normal
Sex
Male
Female
APOE e4 genotype
Carrier
Non−carrier
4y greater education
2 add'l CMC
Dementia to death
0.2
0.3
0.5
0.6
0.8
1.0
1.5
2.0
3.0
5.0
6.0
9.0
Hazard ratio (95% confidence interval)





Males have higher hazard of dementia, equal lifetime risk, smaller sojourn years.

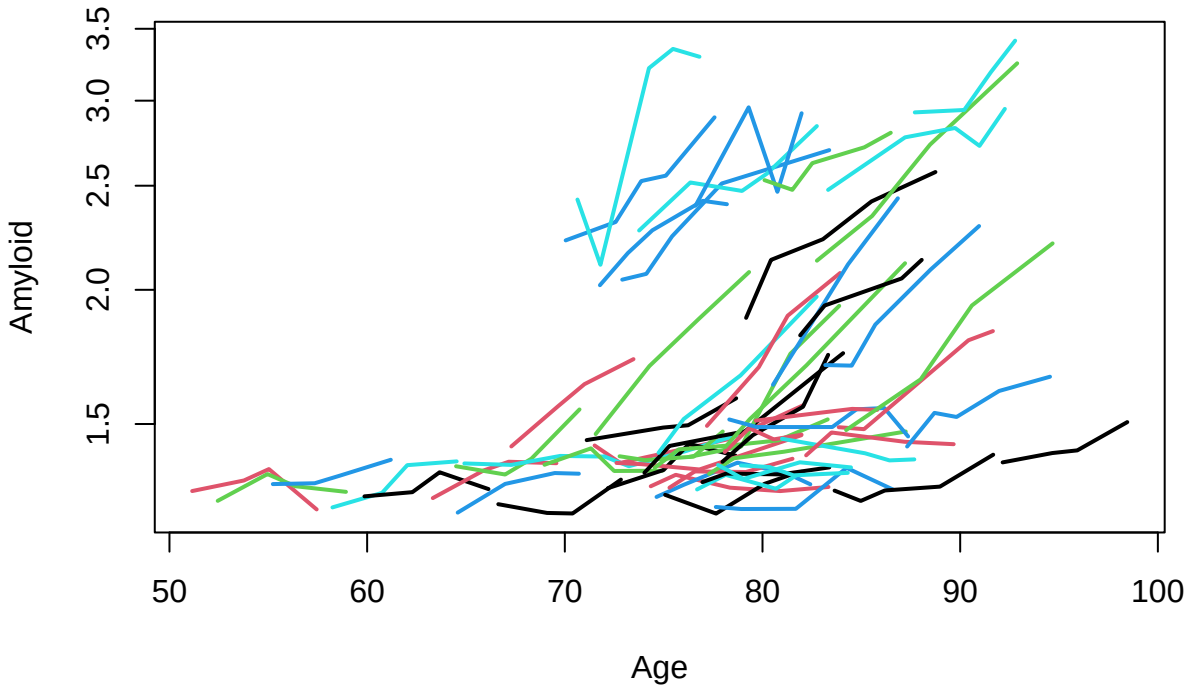
There are a lot of choices in this fit. * The amlyoid level is treated as categorical, to deal with missing. *

An APOE by sex interaction is known from the literature, we sort of have to add it. The amyloid by sex interaction was also significant. * Age scale makes sense * There is an enrollment effect on the CU:death transition. But it doesn't work for the CU:dementia one, due to the 15 month visit interval. We don't expect an enrollment time effect on dementia:death. * This fit used initial amyloid level and intial CMC, which I later realized was worrisome.

Time dependent covariates and age scale

Some rules that I have stated over the years * TD covariates are very useful in Cox models * Survival curves + time dependent covariates doesn't make sense * Use landmark curves * Age scale is superior to entry scale for many problems * Multi-state models are preferred * In a mulit-state model, both HR and absolute risk are needed

For the MCSA paper we have time dependent covariates that *will* rise for a lot of people (amyloid and CMC), age scale makes the most sense, absolute risk is critical. But: the landmark approach no longer makes sense.



Problem: * consider the risk set for a dementia at age 82.35 * amyloid-at-baseline as a covariate * for some subjects that value is 1 month old, for some 10 years old. * what does the amyloid HR mean?

Say you had a variable X that is going up for everyone over time, and that true risk depends on the *current* value of X. In an entry time analysis, in a risk set at 10 years all the baseline X values are out of date, but perhaps they are still approximately correctly ordered.

The COX PL term can always be recentered

$$\frac{e^{\beta x_i}}{\sum_j e^{\beta x_j}} = \frac{e^{\beta(x_i - c)}}{\sum_j e^{\beta(x_j - c)}}$$

so the HR only depends on differences in X. This back of the envelope justification doesn't work on age scale, we really should use the current values of X.

A second point is the splicing problem. When I create a curve for the expected future state of someone who is currently 65. * at age 80, this will involve subjects enrolled at age 70+ (no one has 20 years of fu) * The future for a low amyloid subject will use the risk of someone who was low amyloid when they were enrolled. The curve will by systematically too good. * We inherit the EKM flaws without noticing.

Poisson approx * divide age into 5 year bins * within each bin compute rates as a function of sex, APOE, amyloid, CMC * for someone age 60 with low amyloid, compute the future state probs * take a weighted average of hazards at each age

To do * multi-state model * additive log hazards?