Multi-state Models and the Survival package

Terry Therneau

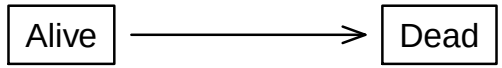
Sept 2019

MAYO CLINIC

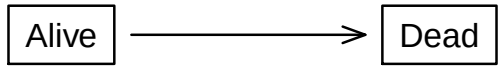
Background

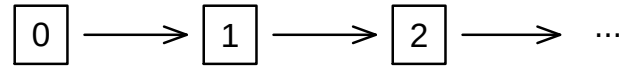
- I am a medical statistician in a tertiary care center
- Most of the questions are time until ...
  - waiting time on the liver transplant list
  - failure of a hip replacement (if ever)
  - clinical trial: chemotherapy to relapse
  - CN $\rightarrow$ MCI $\rightarrow$ Dementia
    none vs. mild vs. moderate/severe neurodegeneration
    none/mild vs. moderate amyloid deposition
- Naturally fell into survial analysis.

MAYO CLINIC

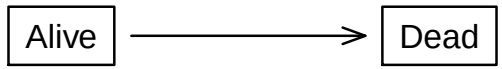


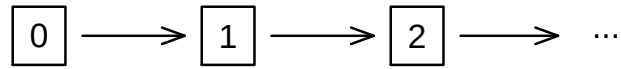
MAYO CLINIC

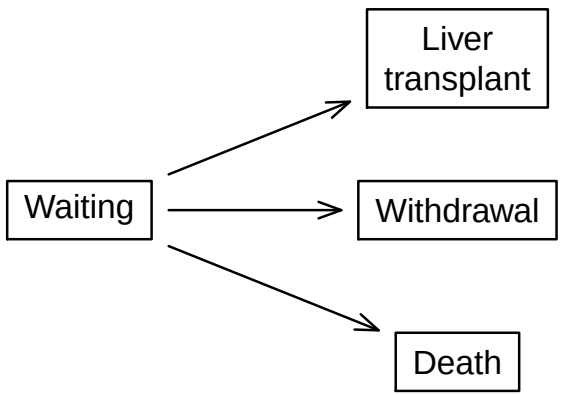




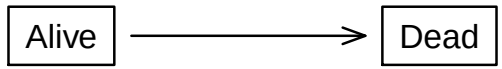
MAYO CLINIC

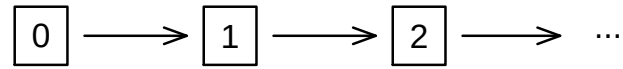


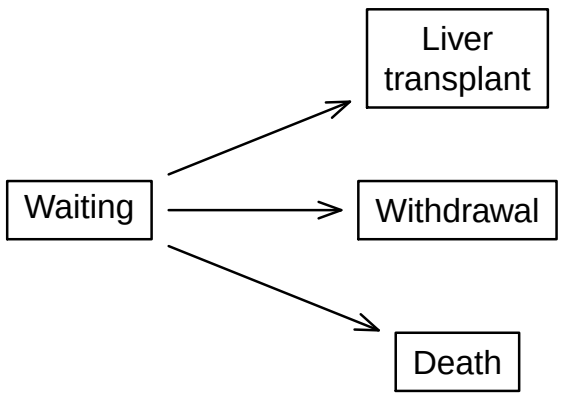


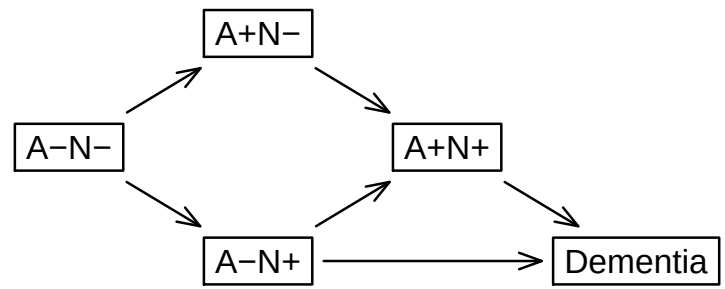


MAYO CLINIC


Alive
Dead


0
1
2
...


Liver transplant
Waiting
Withdrawal
Death


A+N−
A−N−
A+N+
A−N+
Dementia

MAYO CLINIC

Outline

- Three research studies
  - MGUS
  - Leukemia
  - Fatty liver disease
- R code along the way

MAYO CLINIC

Outline

- Three research studies
  - MGUS
  - Leukemia
  - Fatty liver disease
- R code along the way
- Main points
  - Multi-state approach is very useful
  - It ain't that hard

MAYO CLINIC

The survival package

- Does what I need it to do
- Building block for others (666 dependencies on 9/10)

MAYO CLINIC

The survival package

- Does what I need it to do
- Building block for others (666 dependencies on 9/10)
- survival3 went to CRAN on Wed
- Major change/enhancement: multistate

MAYO CLINIC

MGUS

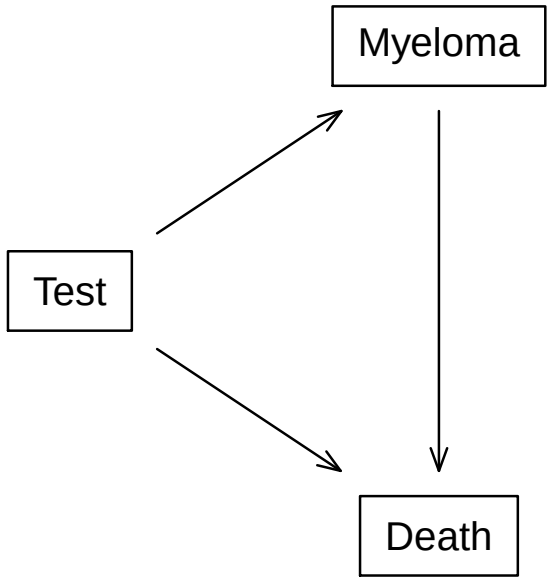
- Observation of Dr Robert Kyle
  - semi-abnormal lab test in normal people (PE)
  - reminiscent of multiple myeloma results
  - are these subjects at risk of MM?
  - note every one he sees, and follow them
- mgus2 data set in R
  - 1384 sequential subjects
  - 5–35 years of follow-up
  - Death, PCM, age, sex, + 3 labs

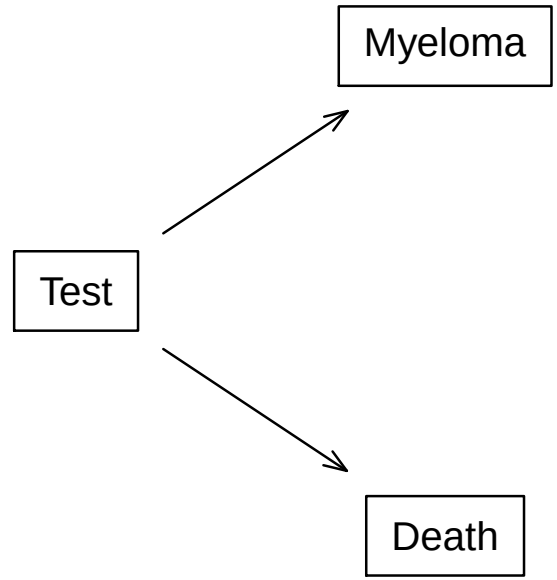
MAYO CLINIC

Data

	id	age	sex	creat	futime	death	ptime	pstat
79	79	96	M	1.8	33	1	33	0
80	80	80	M	1.1	68	1	68	0
81	81	91	F	1.0	21	1	14	1
82	82	71	M	1.1	65	1	65	0
83	83	77	F	0.8	233	1	228	1

MAYO CLINIC





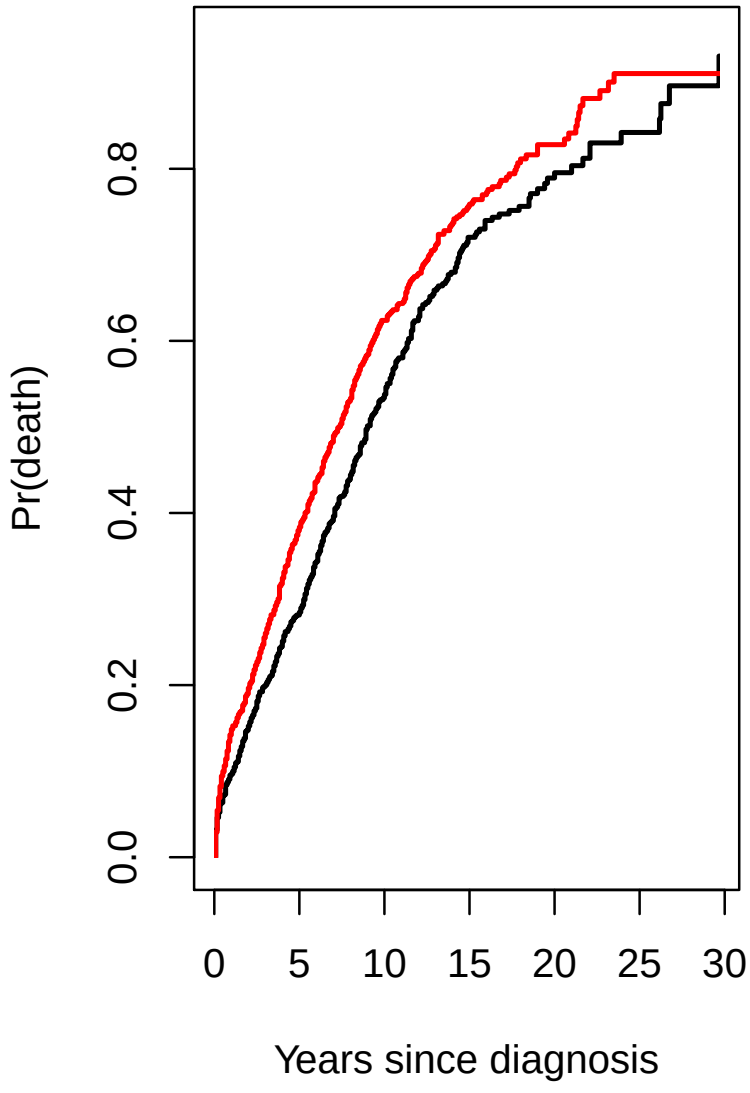
MAYO CLINIC

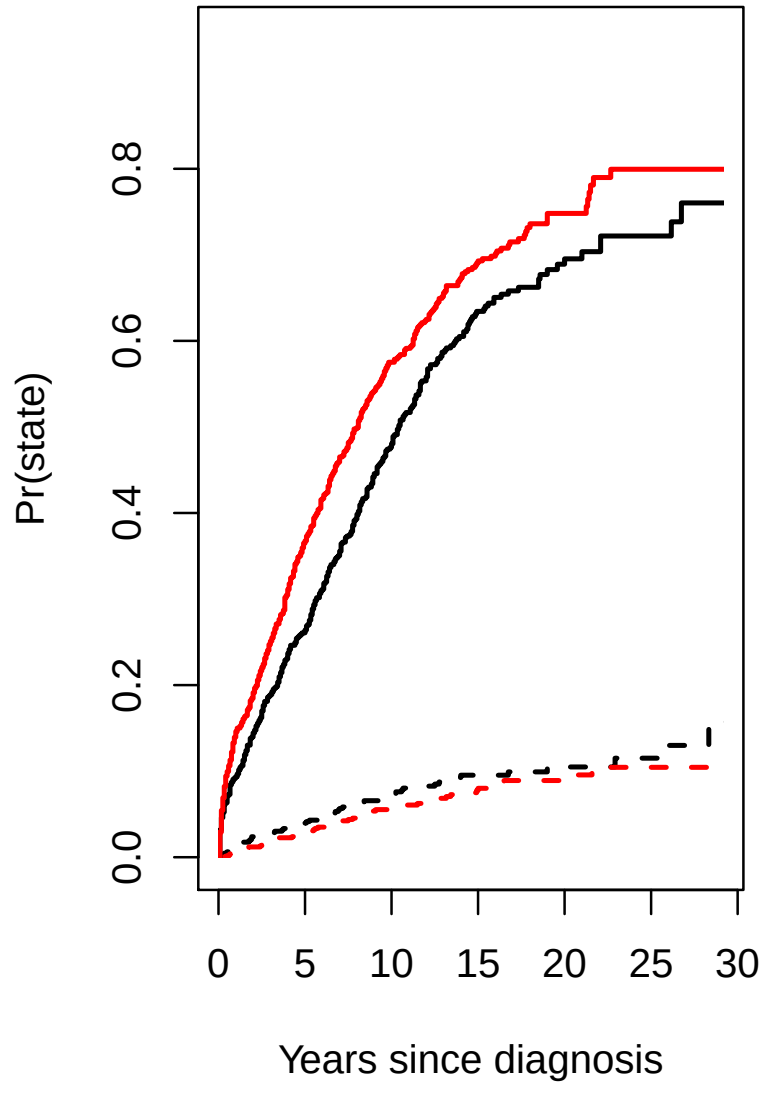
```
    id tstart tstop event age sex  hgb enum
80 79      0    33 Death  96   M  9.6    1
81 80      0    68 Death  80   M 14.1    1
82 81      0    14   PCM  91   F  5.9    1
83 81     14    21 Death  91   F  5.9    2
84 82      0    65 Death  71   M 15.6    1
85 83      0   228   PCM  77   F 12.2    1
86 83    228   233 Death  77   F 12.2    2
87 84      0    78 Death  89   M 14.3    1
```

MAYO CLINIC

```
# Kaplan-Meier
kmfit <- survfit(Surv(futime, death) ~ sex, data=mgus2)

# Competing risks
crdata <- subset(mdata, enum==1)
crfit <- survfit(Surv(tstop, event) ~ sex, data=crdata)
```

MAYO CLINIC


Pr(death)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
30
Years since diagnosis


Pr(state)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
30
Years since diagnosis

MAYO CLINIC

```
# Kaplan-Meier and Cox model
kmfit <- survfit(Surv(futime, death) ~ sex, data=mgus2)
cfit  <- coxph(Surv(futime, death) ~ sex + age, data=mgus2)

# Competing risks and Cox model
crfit <- survfit(Surv(tstop, event) ~ sex, data=crdata)
cfit2 <- coxph(Surv(tstop, event) ~ sex + age10, data=crdata, id=id)
```

MAYO CLINIC

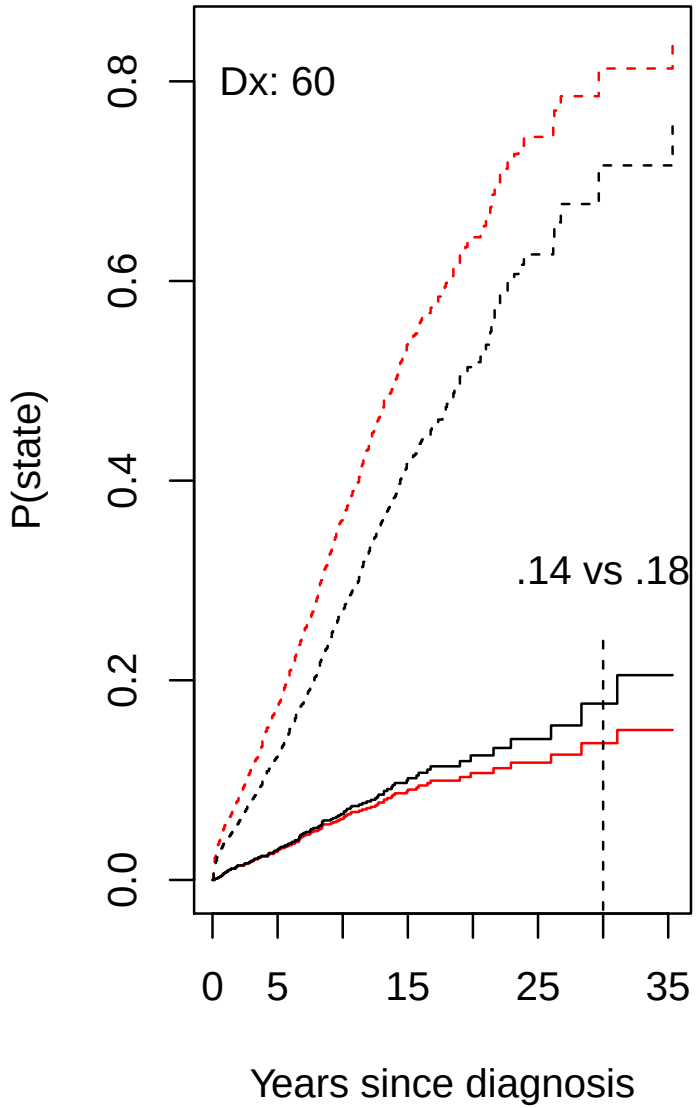
```
Call:
coxph(formula = Surv(tstop, event) ~ sex + age10, data = crdata,
    id = id)


1:2        coef exp(coef) se(coef) robust se    z    p
  sexM   -0.025     0.975    0.188     0.189 -0.1 0.89
  age10   0.130     1.139    0.083     0.067  1.9 0.05


1:3        coef exp(coef) se(coef) robust se    z       p
  sexM   0.393     1.482    0.070     0.066  5.9  3e-09
  age10 0.648     1.912    0.036     0.037 17.6 <2e-16

 States:  1= (s0), 2= PCM, 3= Death

Likelihood ratio test=390  on 4 df, p=<2e-16
n= 1384, number of events= 975
```

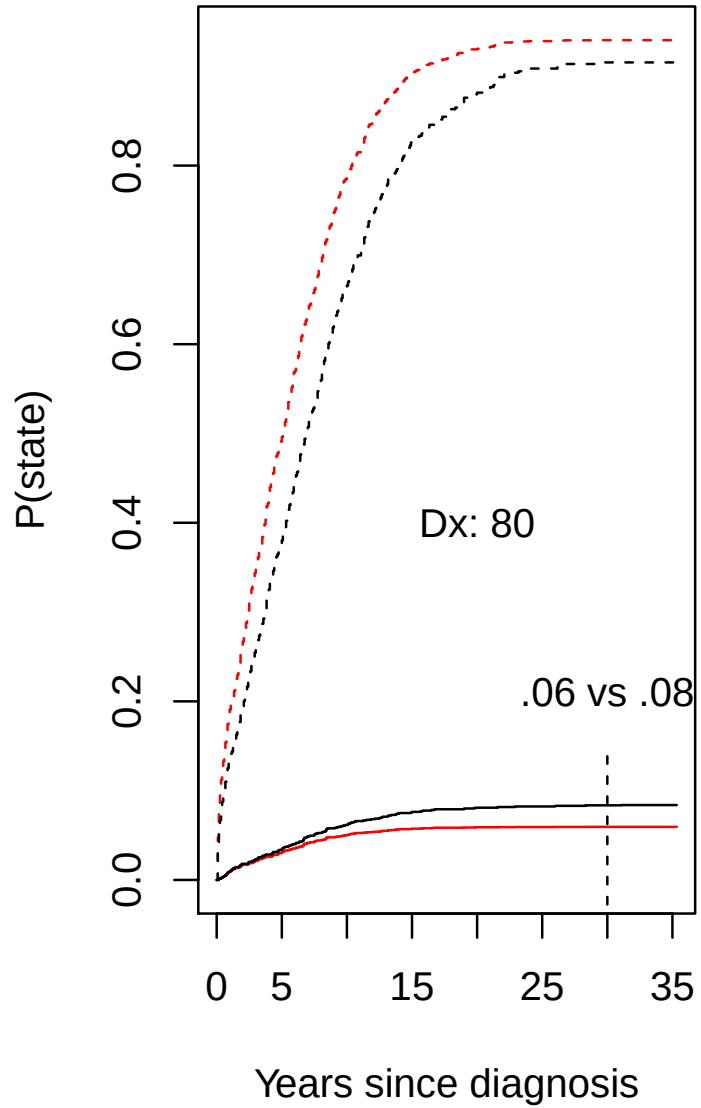
MGUS

- Sex has absolutely no effect on the instantaneous risk of PCM
- Females *do* have an increased lifetime risk of PCM, however.
  - The malignancy rate is 1% per year
  - Females live longer

```
dummy <- expand.grid(sex= c("M", "F"), age10 = c(6, 8))
sfit <- survfit(cfit2, newdata=dummy)   # dimension (4, 3)
plot(sfit[1:2,])
plot(sfit[3:4,])
```

MAYO CLINIC

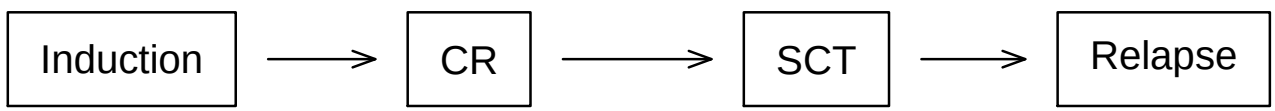

Dx: 60
.14 vs .18
P(state)
0.0
0.2
0.4
0.6
0.8
0
5
15
25
35
Years since diagnosis


Dx: 80
.06 vs .08
P(state)
0.0
0.2
0.4
0.6
0.8
0
5
15
25
35
Years since diagnosis

MAYO CLINIC

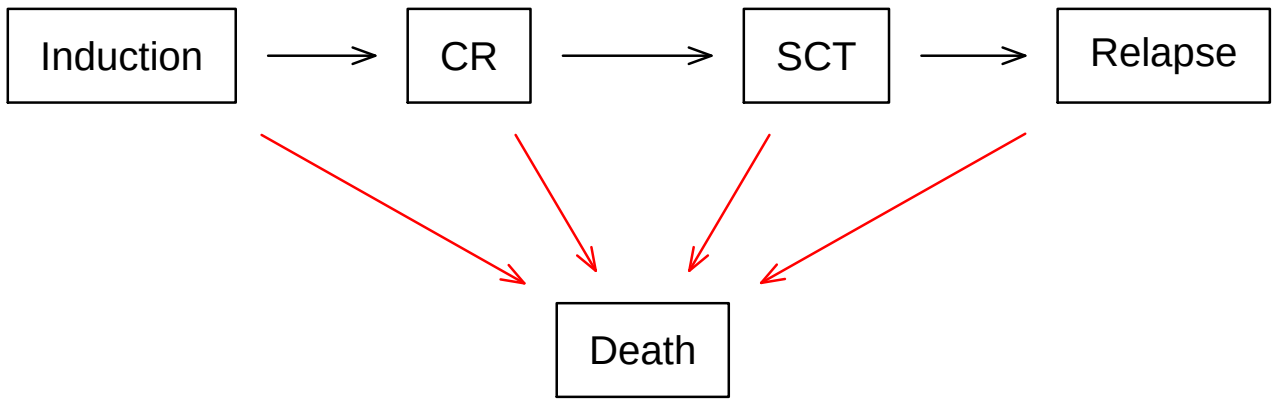
Acute Myeloid Leukemia

- Clinical trial of two induction therapies
- Surprising interaction in the data
- "Can we sort this out?"



MAYO CLINIC

Acute Myeloid Leukemia



MAYO CLINIC

	id	trt	tstart	tstop	event
1	1	B	0	44	CR
2	1	B	44	113	relapse
3	1	B	113	235	death
4	2	A	0	200	SCT
5	2	A	200	286	death
6	3	A	0	38	CR
7	3	A	38	1983	none

MAYO CLINIC

Data

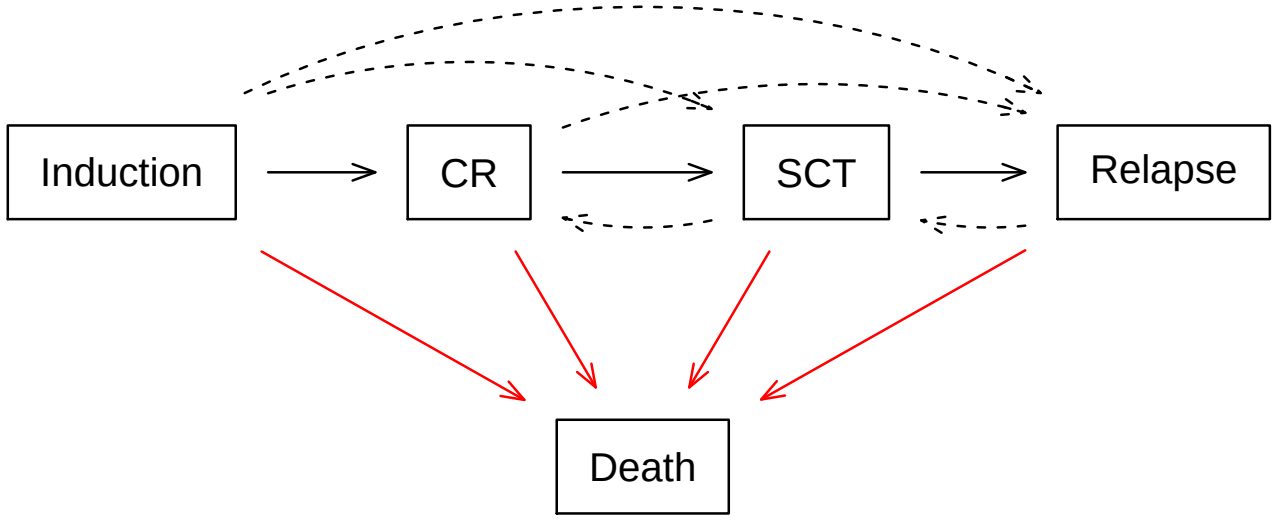
- `event` is a multi-level factor variable.
  - The first level must correspond to "no event at this time"
  - Otherwise unrestricted.
- An `id` variable identifies multiple rows per subject.
- Consistent
  - If at risk, you should be some state: (0,50,CR) (90,210,SCT)
  - But only one place at a time: (0,50,CR) (30,210,SCT)

MAYO CLINIC

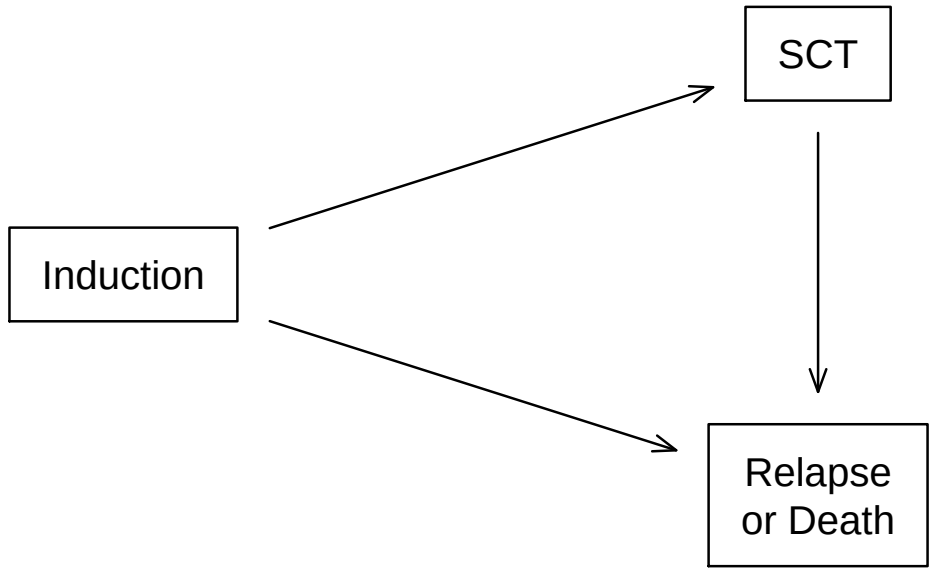
Data check

```
check <- survcheck(Surv(tstart, tstop, event) ~ 1,
                   data=mdata, id=id)
check$transitions
         to
from        CR SCT relapse death (censored)
  (s0)     443 106      13    55         29
  CR         0 159     168    17        110
  SCT       11   0      45   149        158
  relapse    0  99       0    99         28
  death      0   0       0     0          0
#
check$flag
 overlap       gap      jump teleport
       0         0         0        0
```

MAYO CLINIC



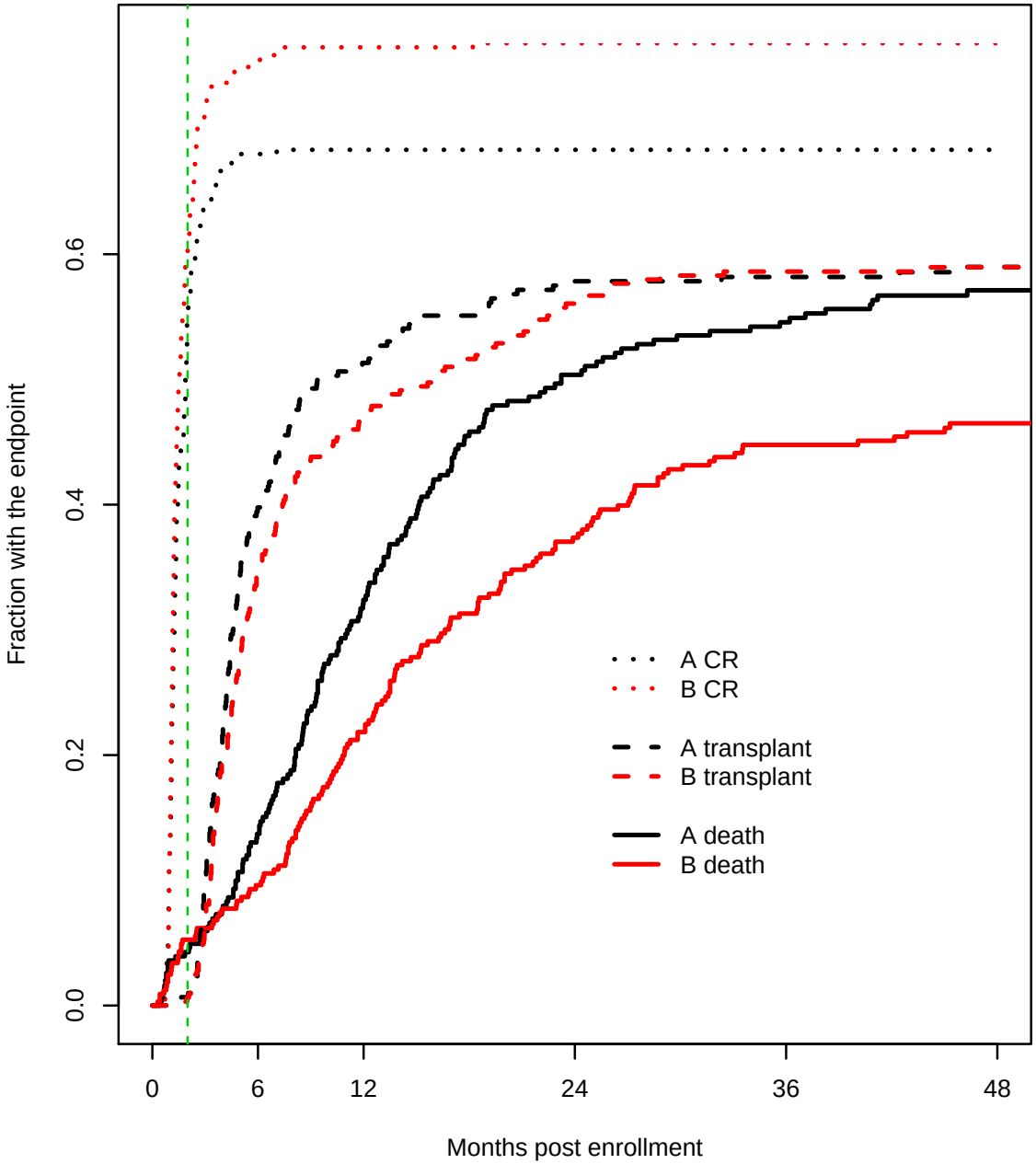
MAYO CLINIC

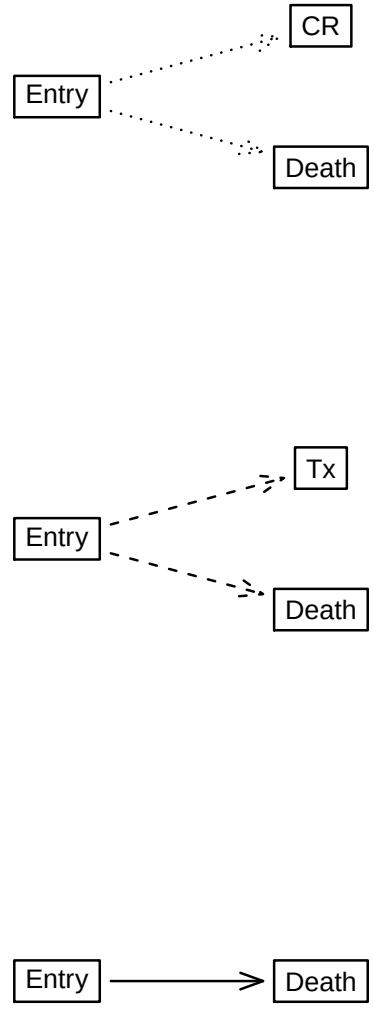


MAYO CLINIC

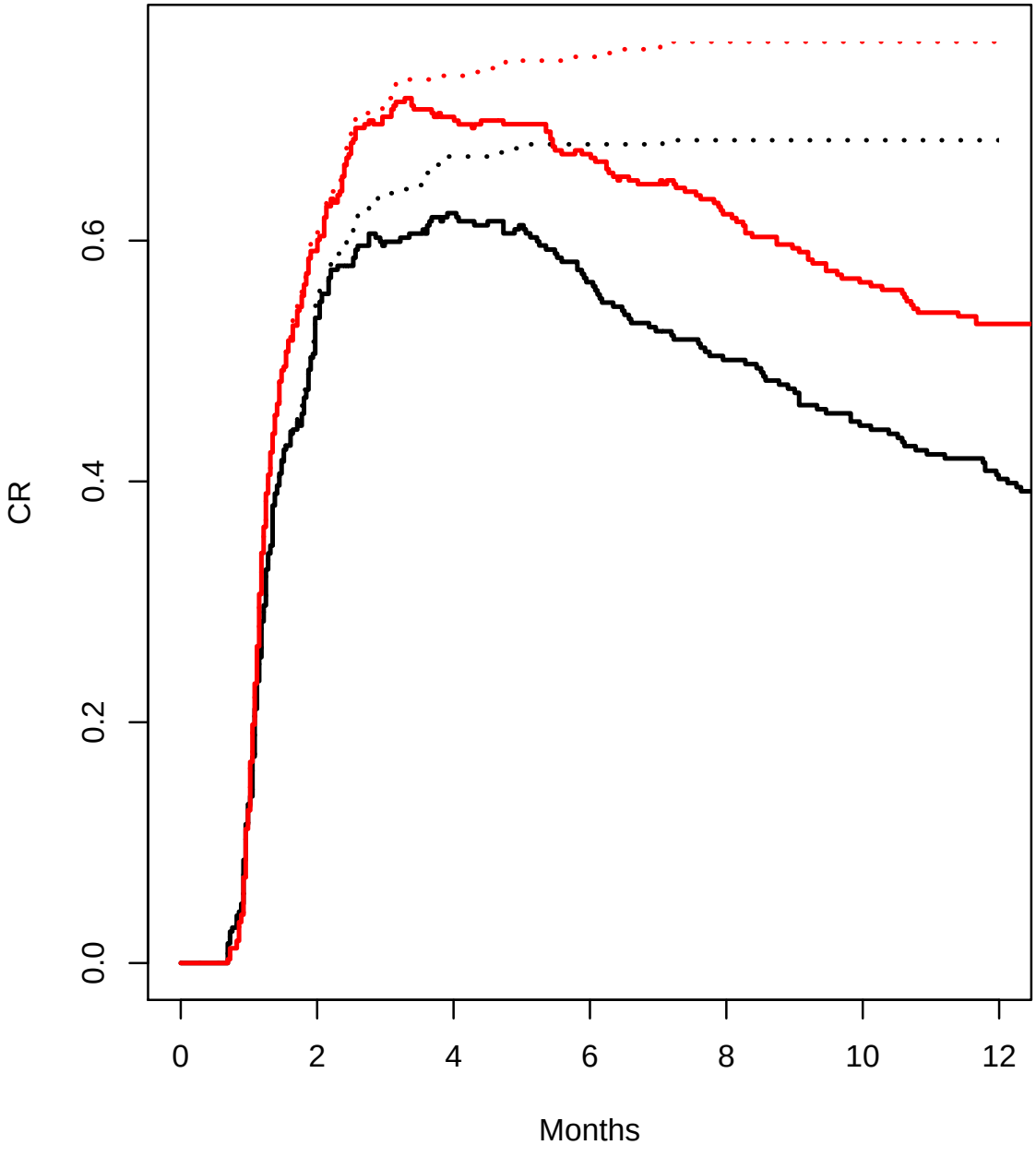
	id	trt	tstart	tstop	event	e2
1	1	B	0	44	CR	none
2	1	B	44	113	relapse	Fail
3	1	B	113	235	death	none
4	2	A	0	200	SCT	SCT
5	2	A	200	286	death	Fail
6	3	A	0	38	CR	none
7	3	A	38	1983	none	none

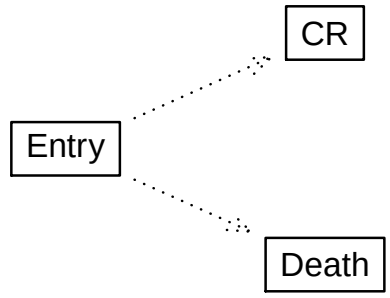
MAYO CLINIC

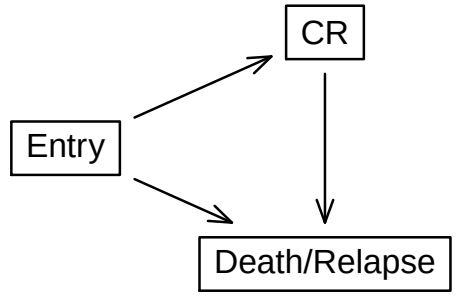

Fraction with the endpoint
0.0
0.2
0.4
0.6
0
6
12
24
36
48
Months post enrollment
A CR
B CR
A transplant
B transplant
A death
B death


Entry
CR
Death
Entry
Tx
Death
Entry
Death

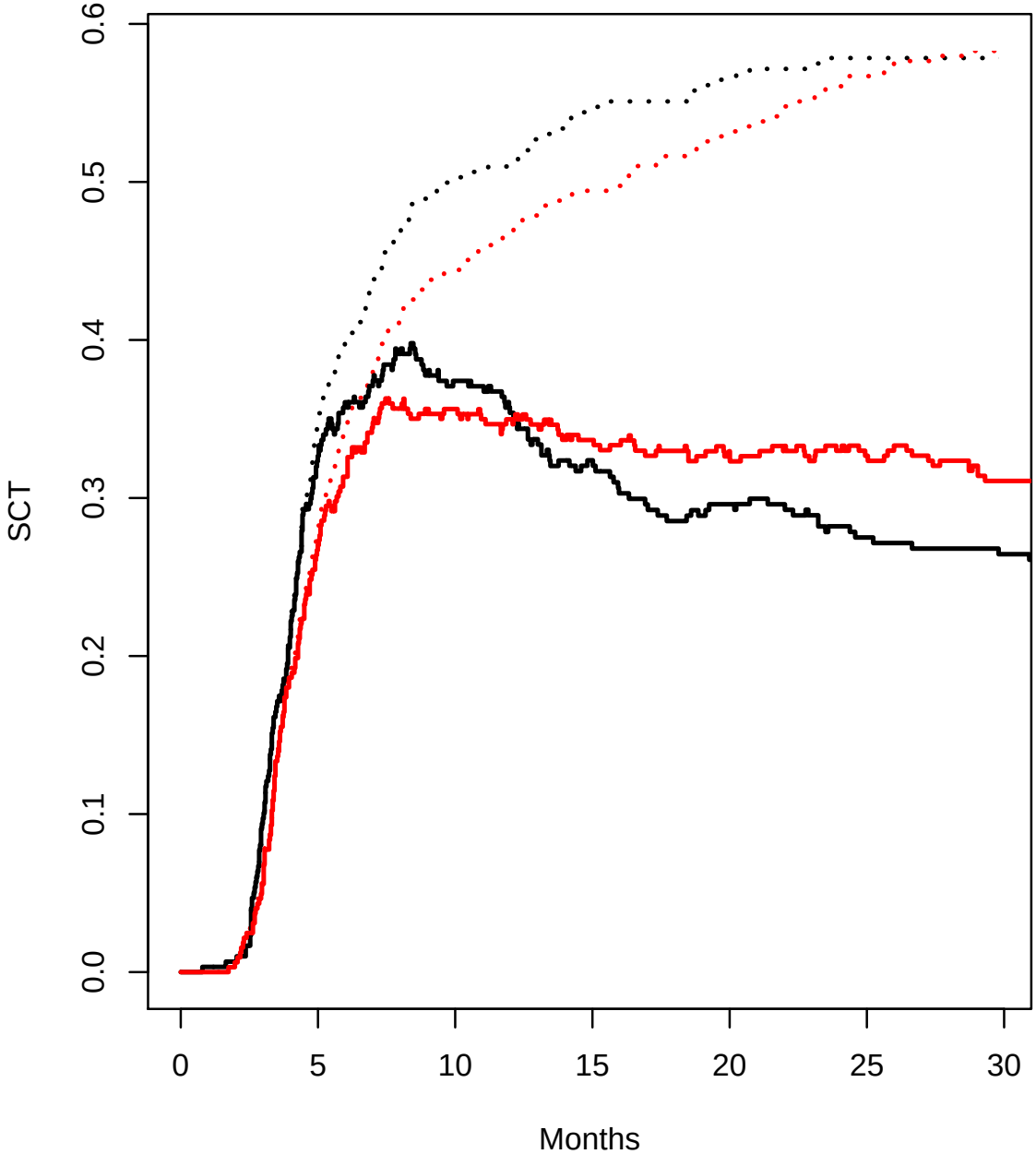
MAYO CLINIC

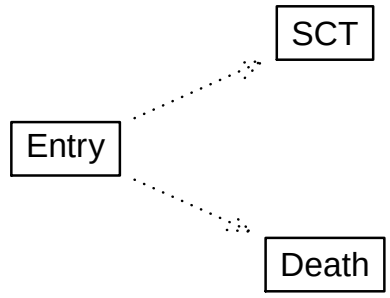

CR
0.0
0.2
0.4
0.6
0
2
4
6
8
10
12
Months

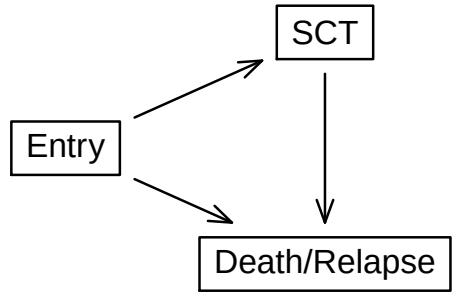

CR
Entry
Death


CR
Entry
Death/Relapse

MAYO CLINIC


SCT
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
15
20
25
30
Months


SCT
Entry
Death


SCT
Entry
Death/Relapse

MAYO CLINIC

Possible questions

- Does the time from CR to progression/death differ?
- Does the rate of SCT after CR differ?
- How do all of these interact with sex?

MAYO CLINIC

Possible questions

- Does the time from CR to progression/death differ?
- Does the rate of SCT after CR differ?
- How do all of these interact with sex?
- Why would you settle for a single (boring) KM?

MAYO CLINIC

AJ estimator

- The survfit routine caculates the Aalen-Johansen estimate of $p(t)$
  $p_j(t)$ =Pr(in state j at time t)
- $\sum_j p_j(t) = 1$
- IJ variance
- Kaplan-Meier = special case of AJ
- Cumulative incidence = special case of the AJ

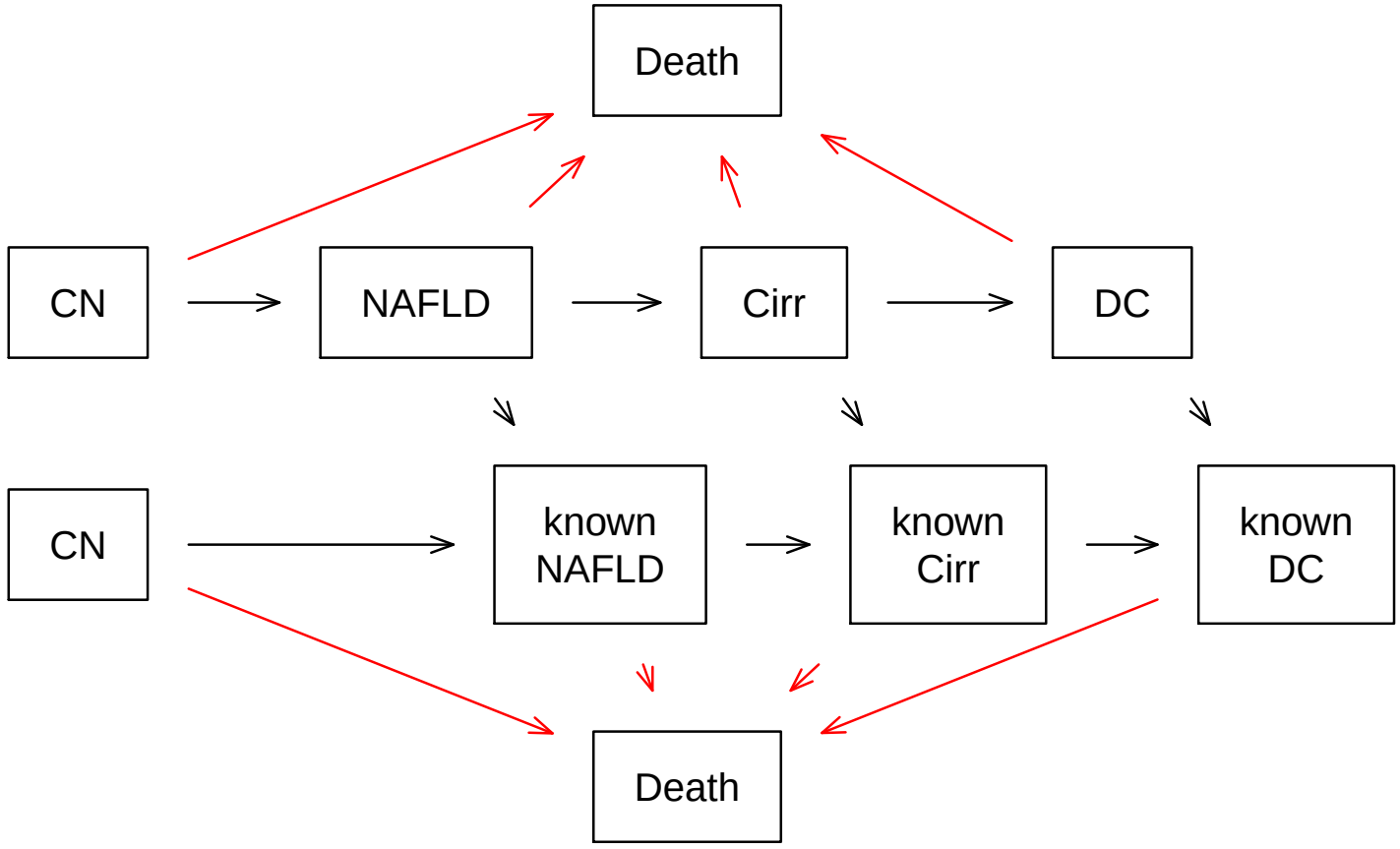
MAYO CLINIC

Fatty liver disease (NAFLD)

- Definition
  - presence of $> 5\%$ fat in the liver (steatosis)
  - absence of other indications (alcohol or certain medications)
  - no other liver disease
- Why would you care?
  - responsible for $1/3$ of liver transplants
  - prevalence is growing with obesity

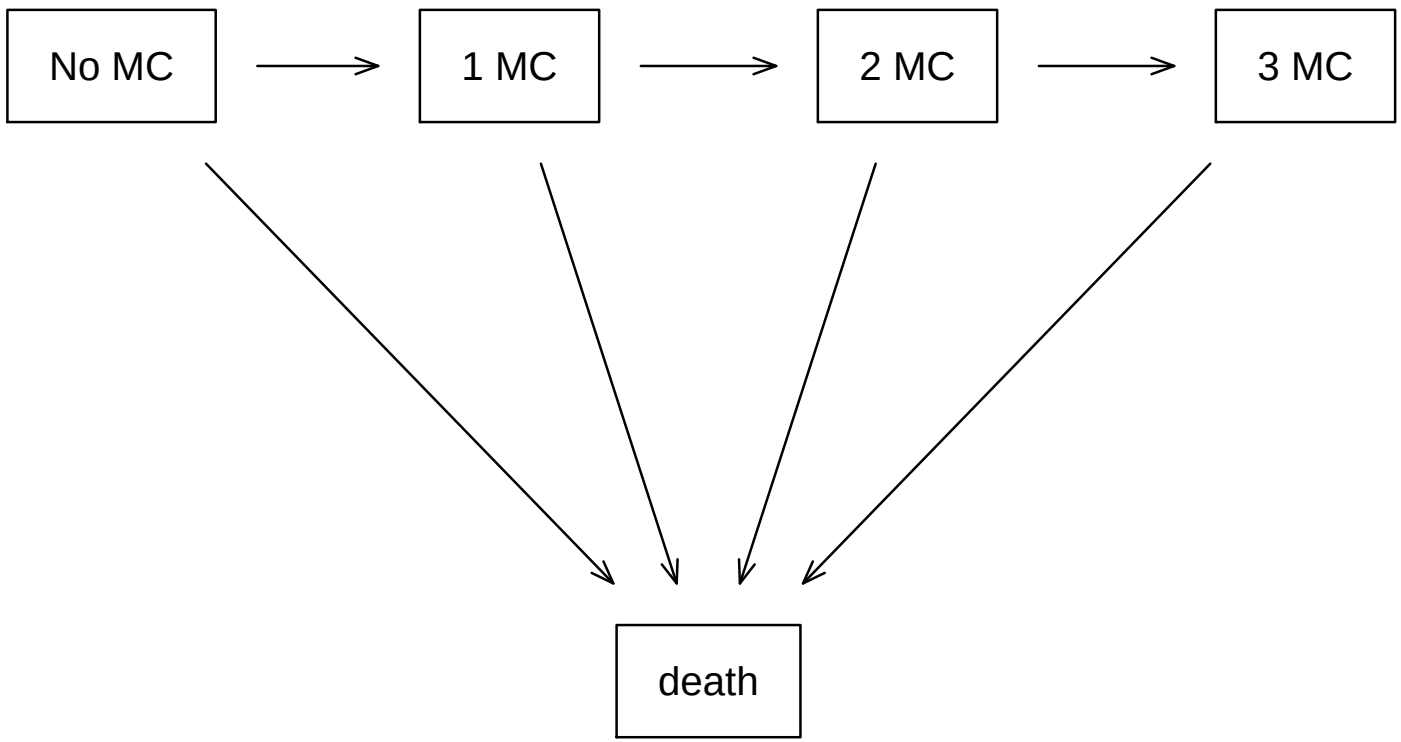
"Nonalcoholic fatty liver disease, metabolic syndrome, and the fight that will define clinical practice for a generation of hepatologists." Tapper and Loomba, Hepatology 2018.

MAYO CLINIC


Death
CN
NAFLD
Cirr
DC
CN
known
NAFLD
known
Cirr
known
DC
Death

MAYO CLINIC

- NAFLD as a covariate
- Outcome of 0–3 “metabolic syndrome” comorbidities
  - diabetes
  - dyslipidemia
  - hypertension



MAYO CLINIC

Data

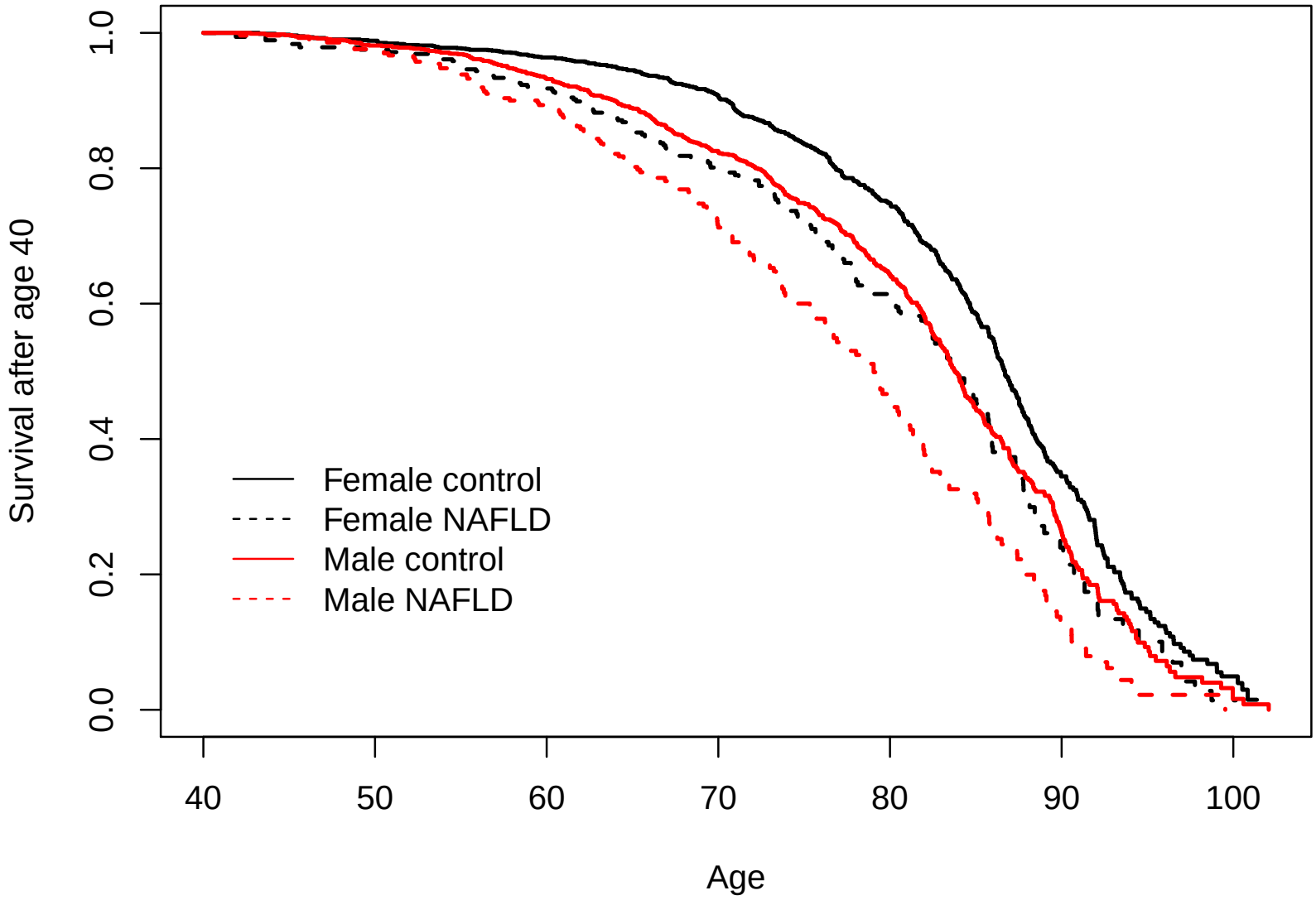
- 3514 NAFLD cases
- 14004 age/sex matched controls

MAYO CLINIC

Data

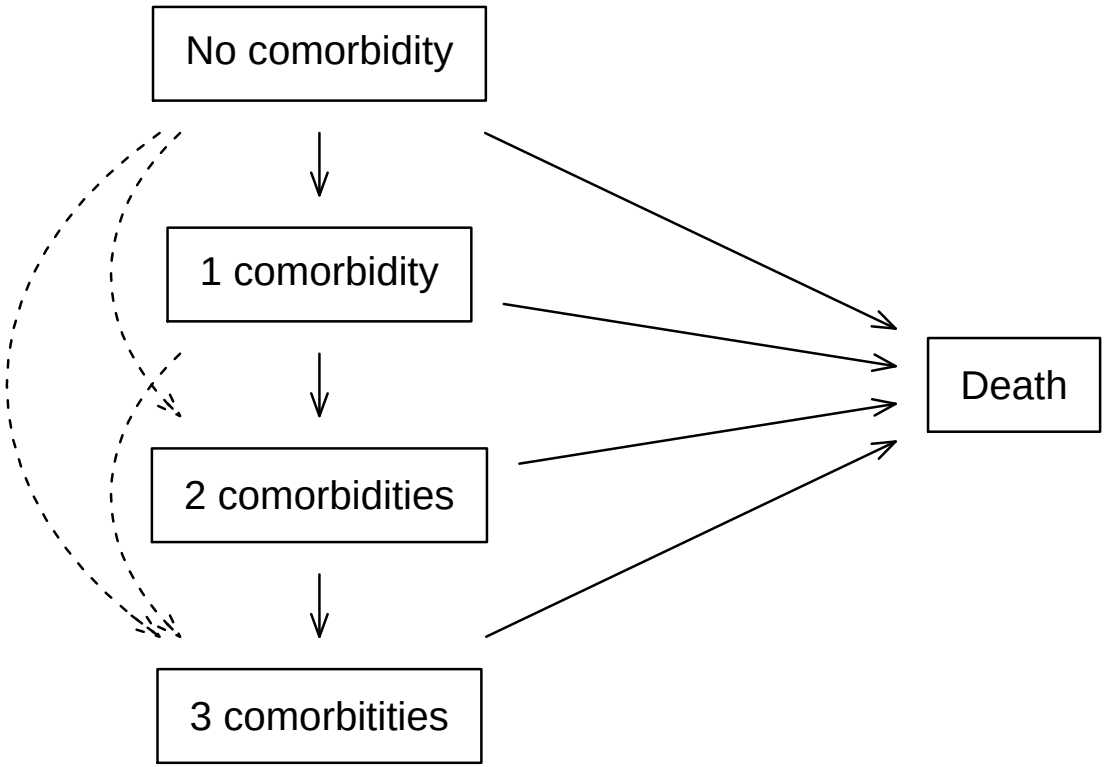
- 3514 NAFLD cases
- 14004 age/sex matched controls
- Use age scale
- "Duration of NAFLD" not possible

MAYO CLINIC


Survival after age 40
1.0
0.8
0.6
0.4
0.2
0.0
40
50
60
70
80
90
100
Age
Female control
Female NAFLD
Male control
Male NAFLD

MAYO CLINIC

Multistate



MAYO CLINIC

```
check <- survcheck(Surv(age1, age2, event) ~ 1, data=data1,
          istate=cstate, id = id)
check$transitions
       to
from      1mc  2mc  3mc death (censored)
  0mc    1829   70    4   263        5705
  1mc       0 1843   28   243        4567
  2mc       0    0 1048   417        3687
  3mc       0    0    0   441        2220
  death     0    0    0     0           0
```

MAYO CLINIC

Analysis

- Hazard ratios
- Absolute risk
  - current state $p(t)$
  - E(time in state)
  - E(number of visits to each state)
    lifetime risk
  - E(time in state, per visit)

MAYO CLINIC

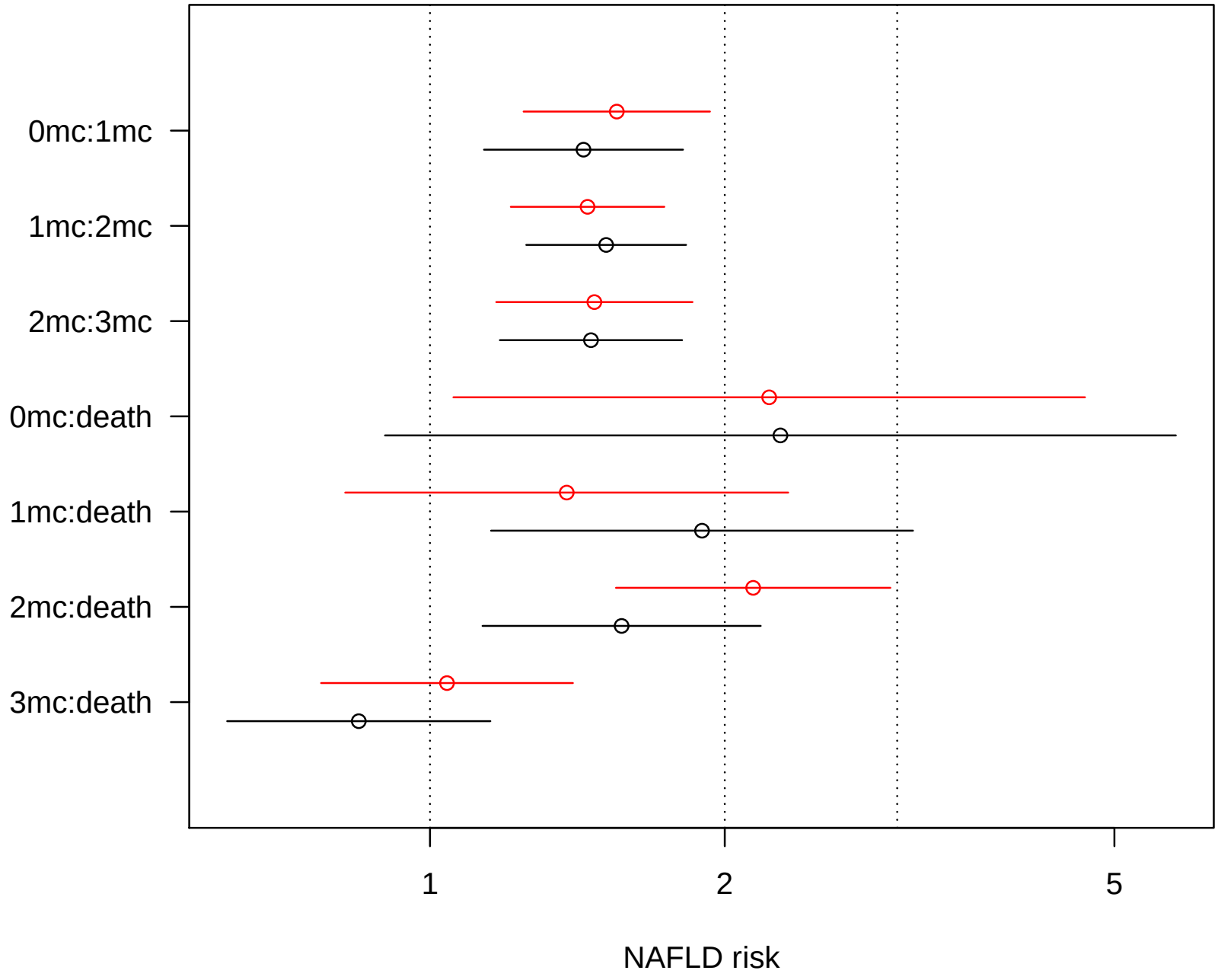
Hazard model

- Default
  - separate baseline for each transition
  - separate coefficients for each transition
- 0mc:3mc has 4 cases, 1:3 only has 28
- Collapse 0:1, 0:2, 0:3 = "worsening"
- x:death as a single baseline + covariates?
- Many choices possible

MAYO CLINIC

Cox fits

```
cform <- list(Surv(age1, age2, event) ~ nafld + bmi,
              "0mc":c("1mc", "2mc", "3mc") ~ nafld +bmi/ common,
              "1mc":c("2mc", "3mc")        ~ nafld +bmi/ common)

coxm <- coxph(cform, istate=cstate, id= id,
                 data1, subset= (male==1))
coxf <- coxph(cform, istate=cstate, id = id,
                data1, subset= (male==0))
coxm$cmap
           1:2 1:3 2:3 1:4 2:4 3:4 1:5 2:5 3:5 4:5
(Baseline)   1   2   3   4   5   6   7   8   9  10
nafld        1   1   3   1   3   5   7   9  11  13
bmi          2   2   4   2   4   6   8  10  12  14
j <- unique(coxm$cmap[2,])
temp <- rbind(male=coef(coxm)[j], female= coef(coxf)[j])
colnames(temp) <- c("0:1", "1:2", "2:3", paste(0:3, 'death', sep=':'))
round(exp(temp),2)
        0:1  1:2  2:3 0:death 1:death 2:death 3:death
male   1.55 1.45 1.47    2.22    1.38    2.14    1.04
female 1.43 1.51 1.46    2.28    1.90    1.57    0.85
```


0mc:1mc
1mc:2mc
2mc:3mc
0mc:death
1mc:death
2mc:death
3mc:death
1
2
5
NAFLD risk

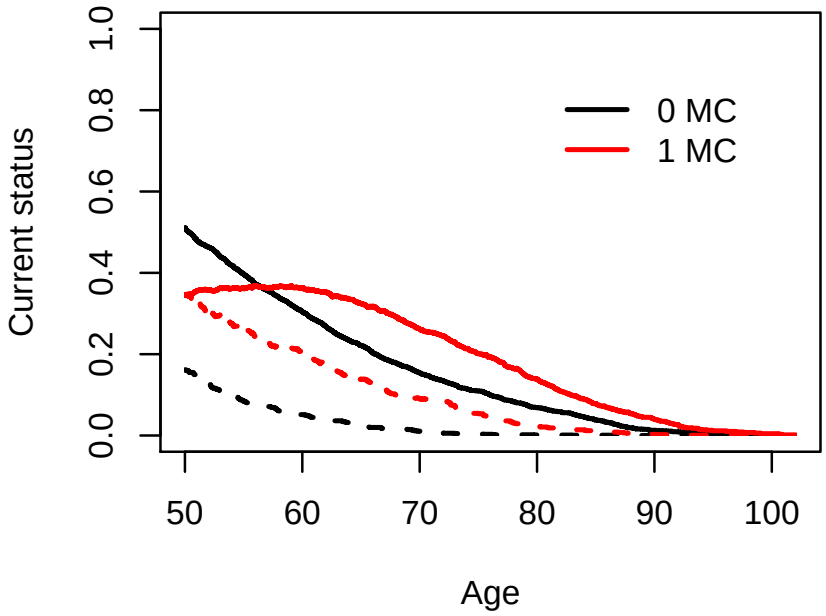
MAYO CLINIC

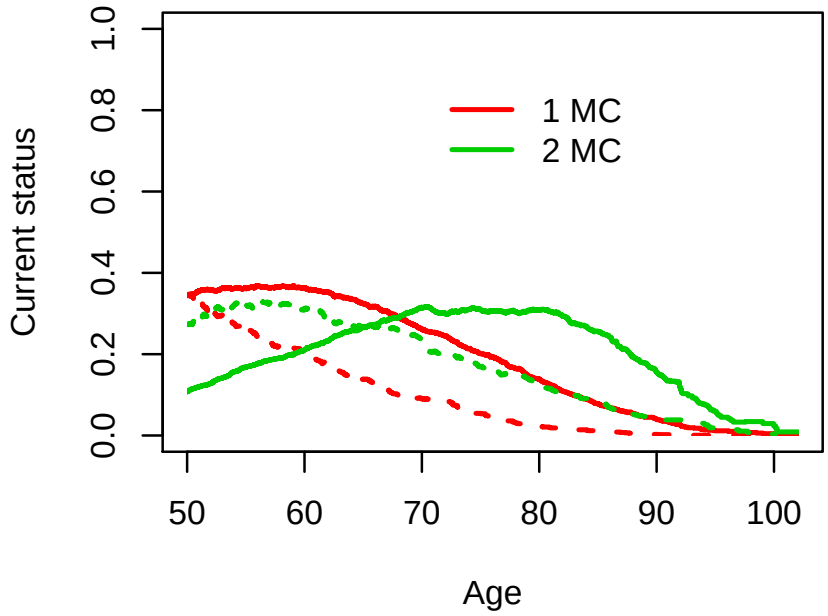
State at enrollment

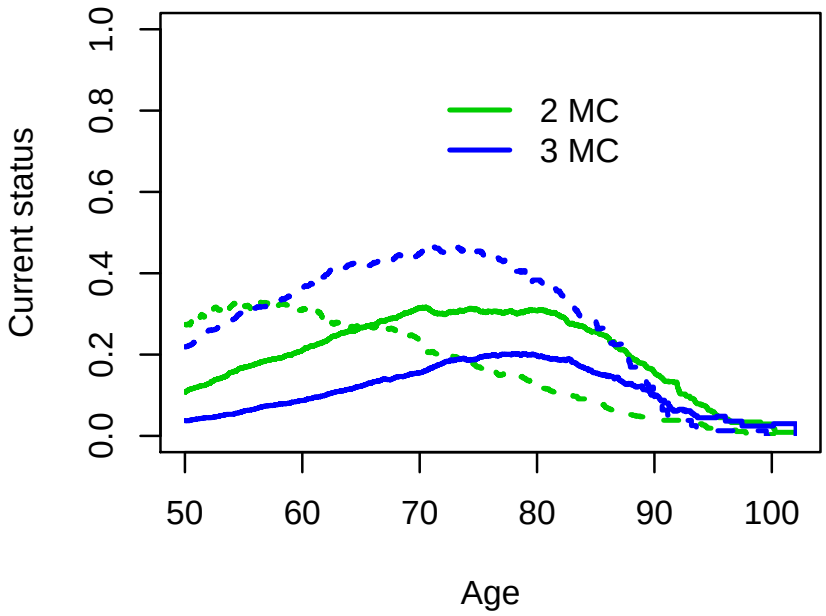
	cstate			
nafld0	0mc	1mc	2mc	3mc
0	51	27	16	7
1	22	31	28	19

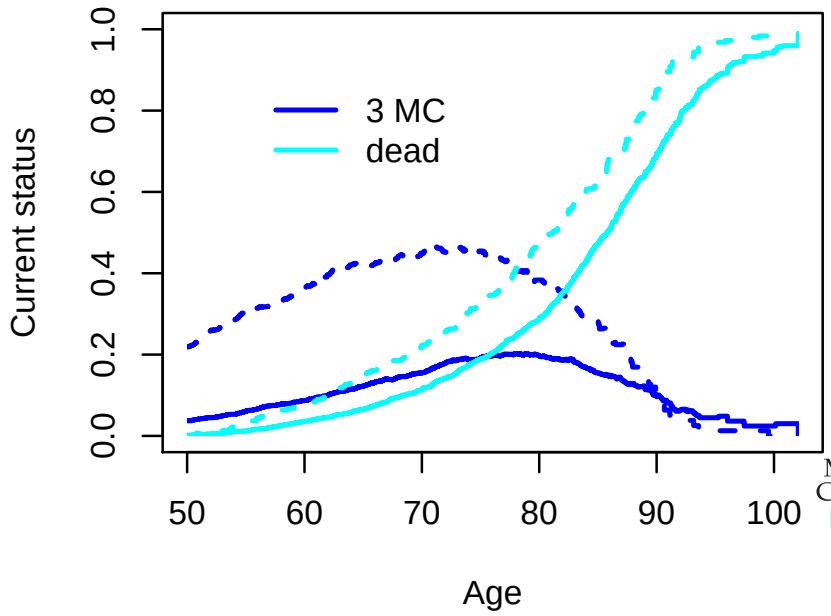
MAYO CLINIC

Current status as a function of age







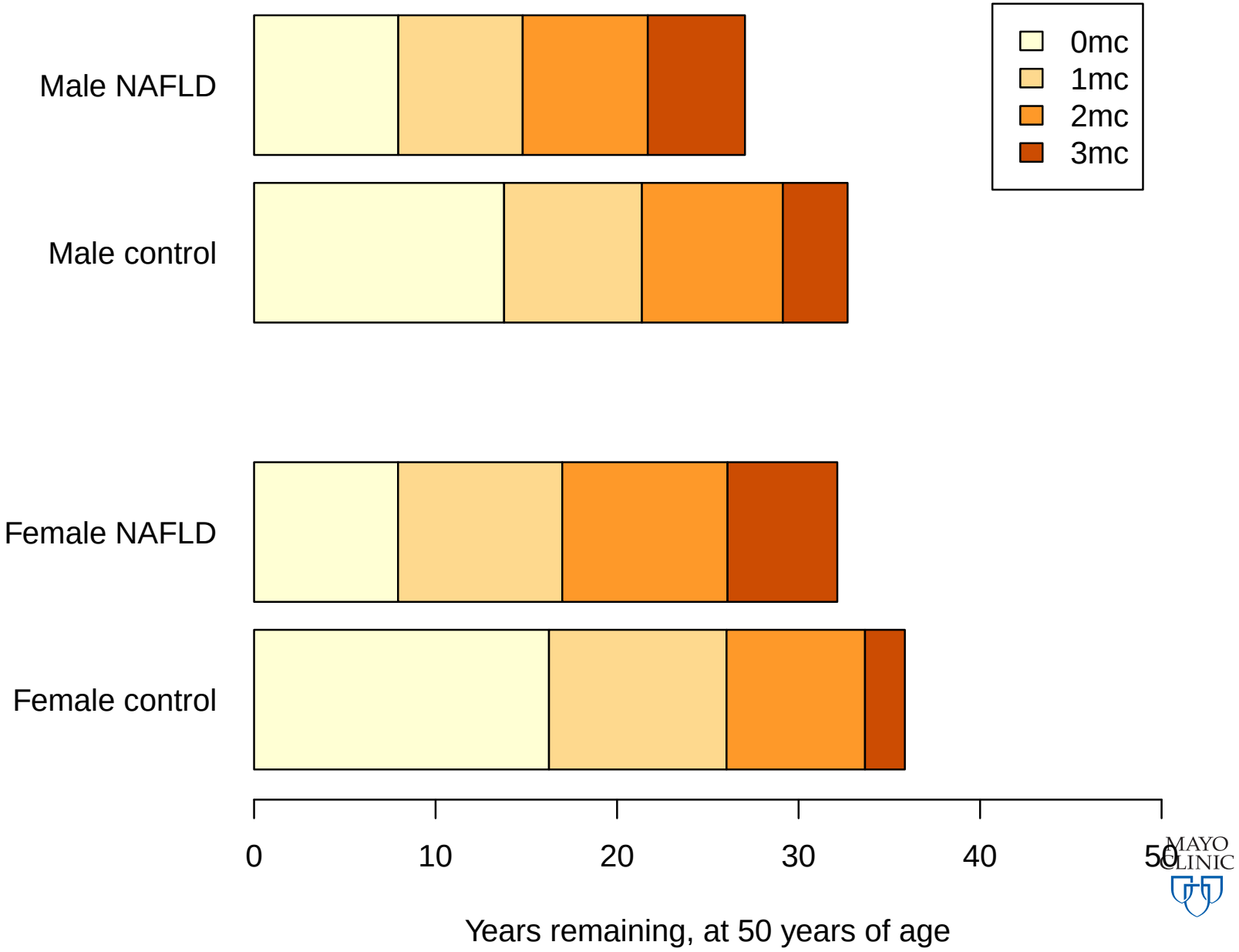


MAYO CLINIC

Expected future for a 50 year old with no comorbidities

```
s50a <-  survfit(Surv(age1, age2, event) ~ nafld0, data=data1,
                 istate=cstate,
                 id=id, start.time=50)

s50b <- survfit(Surv(age1, age2, event) ~ nafld0, data=data1,
                istate=cstate,
                id=id, start.time=50,
                p0=c(1,0,0,0,0))

print(s50b, rmean=100, digits=2)
Call: survfit(formula = Surv(age1, age2, event) ~ nafld0, data = data1,
    id = id, istate = cstate, start.time = 50, p0 = c(1, 0, 0,
        0, 0))

                     n nevent rmean*
nafld0=0, 0mc    13309      0   15.2
nafld0=1, 0mc     3283      0    7.8
nafld0=0, 1mc    13309    980    8.9
nafld0=1, 1mc     3283    129    8.2
nafld0=0, 2mc    13309   1143    7.7
nafld0=1, 2mc     3283    327    8.2
nafld0=0, 3mc    13309    575    2.7
```

Time in state



Summary

- Multi-state problems are common
- Rich analysis possibilities
- Tools are getting better
- Just don't get carried away with the wrong things . . .

MAYO CLINIC

Software goals

- Multi-state fits are as easy as `coxph`
- Curves of $P(\text{state} \mid t)$ are as easy as `survfit`
- Secondary summaries (time in state, E(visits to state), . . .)
- Better data tools
  - Rules for data
    - Cannot be in two places at once
    - If at risk, must be someplace
    - Gaps and teleports are viewed with suspicion
    - Don't make the user break the rules
  - Immortal time bias is everywhere.
- These are the tools that I use.

MAYO CLINIC

Software goals

- Multi-state fits are as easy as `coxph`
- Curves of $P(\text{state} \mid t)$ are as easy as `survfit`
- Secondary summaries (time in state, E(visits to state), . . .)
- Better data tools
  - Rules for data
    - Cannot be in two places at once
    - If at risk, must be someplace
    - Gaps and teleports are viewed with suspicion
    - Don't make the user break the rules
  - Immortal time bias is everywhere.
- These are the tools that I use.
- Don't bugger it up

MAYO CLINIC