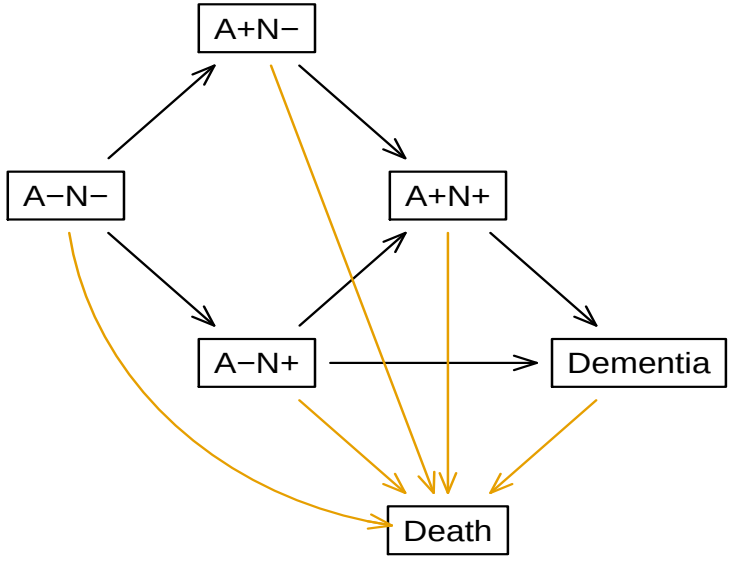
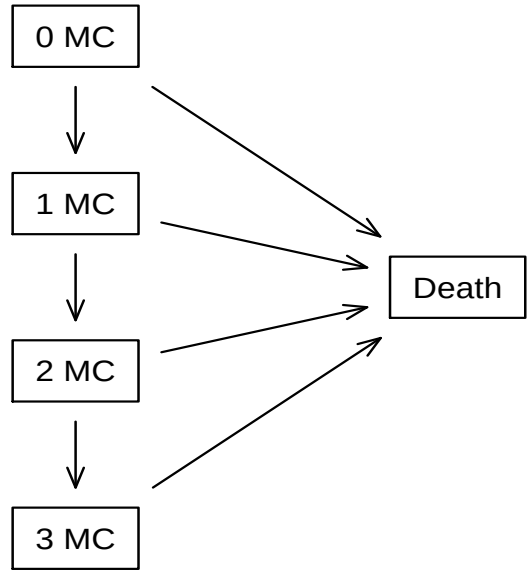
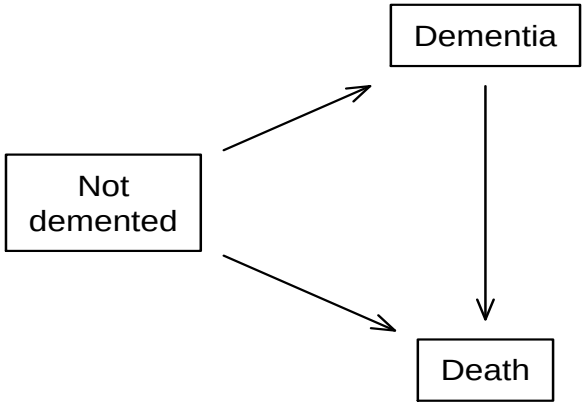
A multi-state model for dementia

Terry Therneau

May 2021


A+N−
A−N−
A+N+
A−N+
Dementia
Death


0 MC
1 MC
2 MC
3 MC
Death


Dementia
Not
demented
Death

Overview

- Multi-state hazard models are very useful.
  - Easy to fit (R survival package)
- Box and arrow diagram
- Transition rates (arrows)
  - Each arrow is an individual Cox model
  - Additive, linear, proportional hazards
  - Non-informative censoring
  - Avoid immortal time bias

Overview

- Multi-state hazard models are very useful.
  - Easy to fit (R survival package)
- Box and arrow diagram
- Transition rates (arrows)
  - Each arrow is an individual Cox model
  - Additive, linear, proportional hazards
  - Non-informative censoring
  - Avoid immortal time bias
- Boxes (absolute risk)
  - Must be dealt with in toto
  - Probability in state s at time t
  - E(number of visits to state s) = lifetime risk
  - E(time in state s)

Overview

- Multi-state hazard models are very useful.
  - Easy to fit (R survival package)
- Box and arrow diagram
- Transition rates (arrows)
  - Each arrow is an individual Cox model
  - Additive, linear, proportional hazards
  - Non-informative censoring
  - Avoid immortal time bias
- Boxes (absolute risk)
  - Must be dealt with in toto
  - Probability in state s at time t
  - E(number of visits to state s) = lifetime risk
  - E(time in state s)
- 80% data, 40% tune individual fits, 20% overall fit

Mayo Clinic Study of Aging

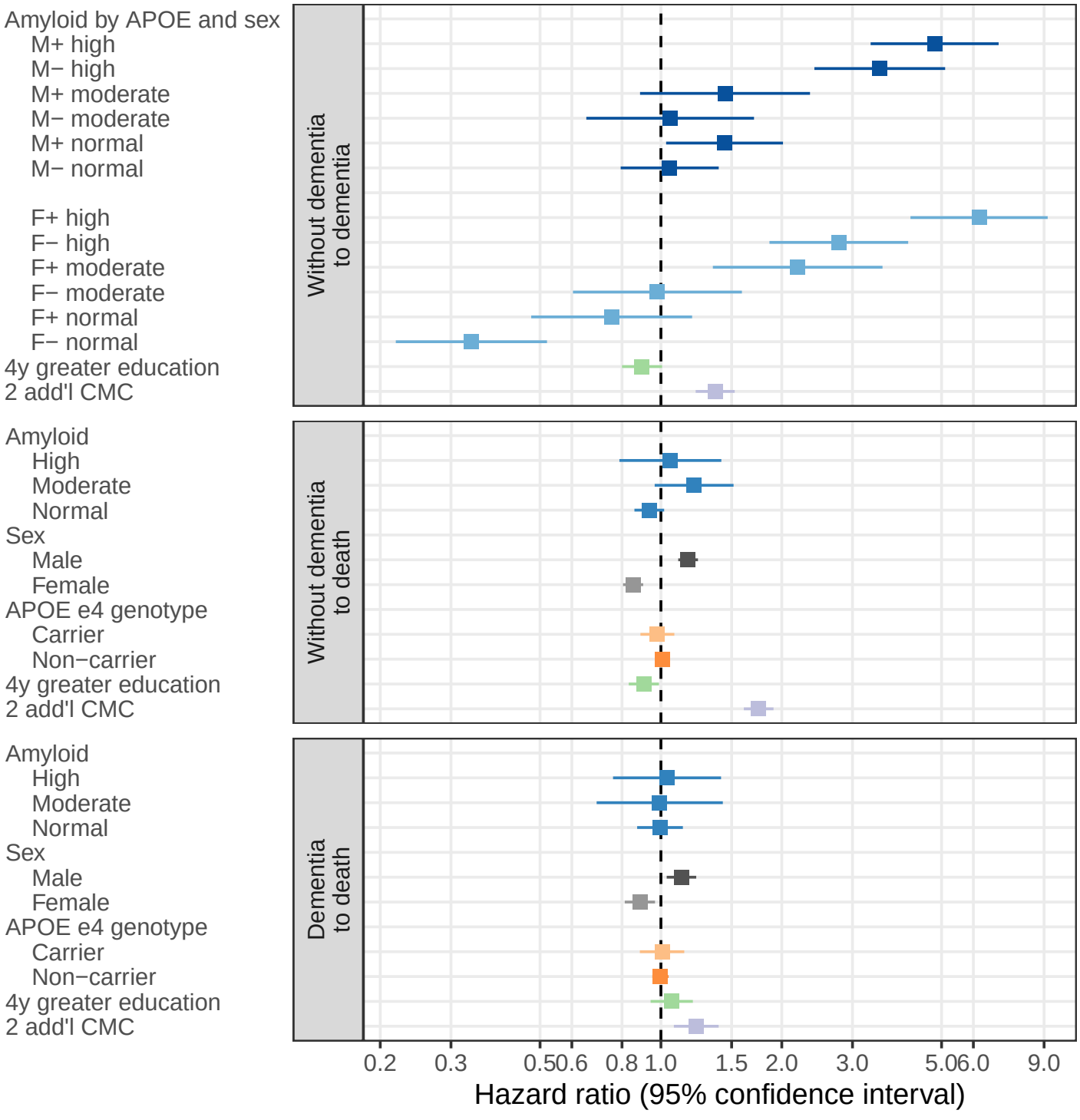
- Population based study in Olmsted County, Minn
- Age and sex stratified random sample
- Scheduled visits every 15 months
- Active set of approx 3000

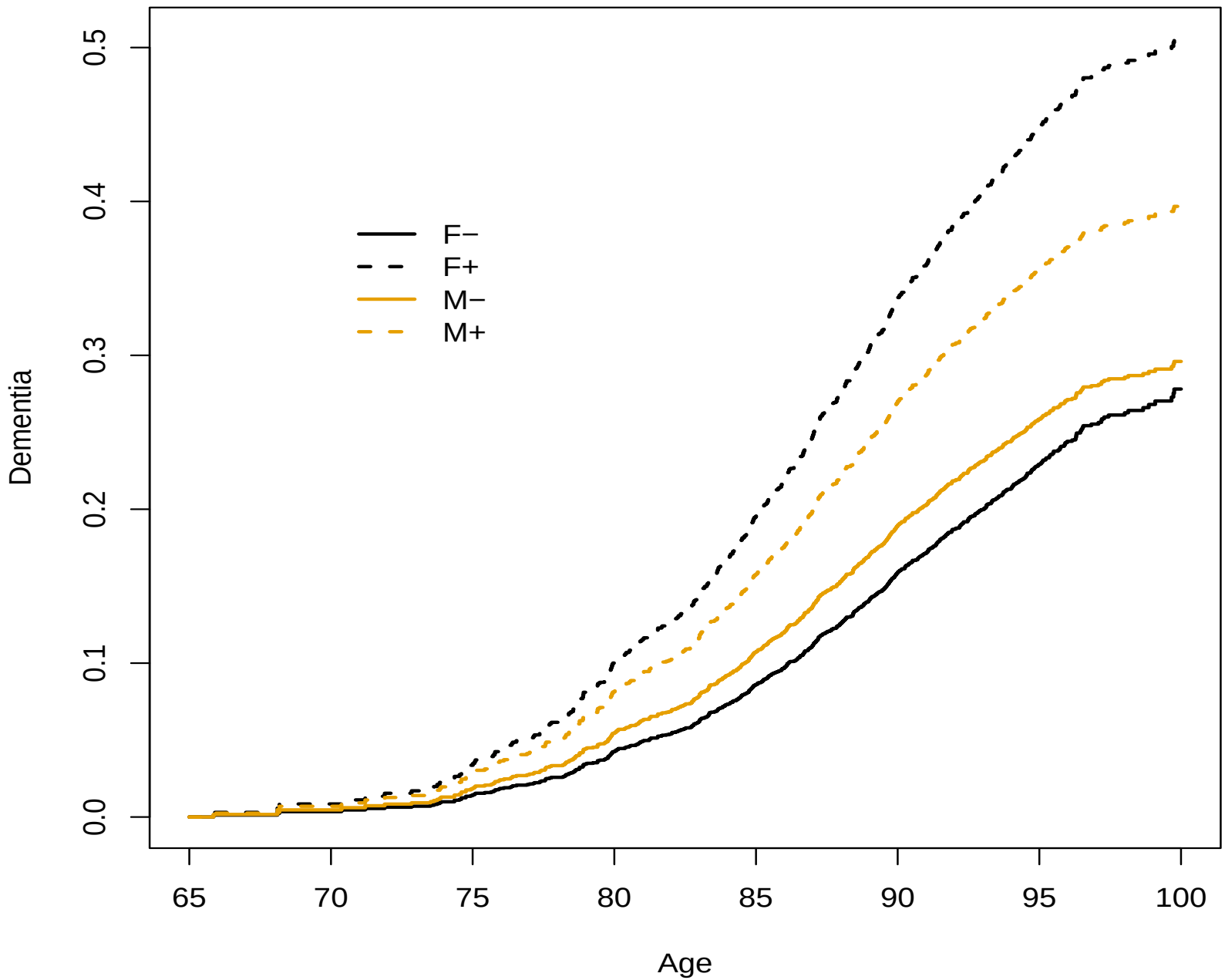
Mayo Clinic Study of Aging

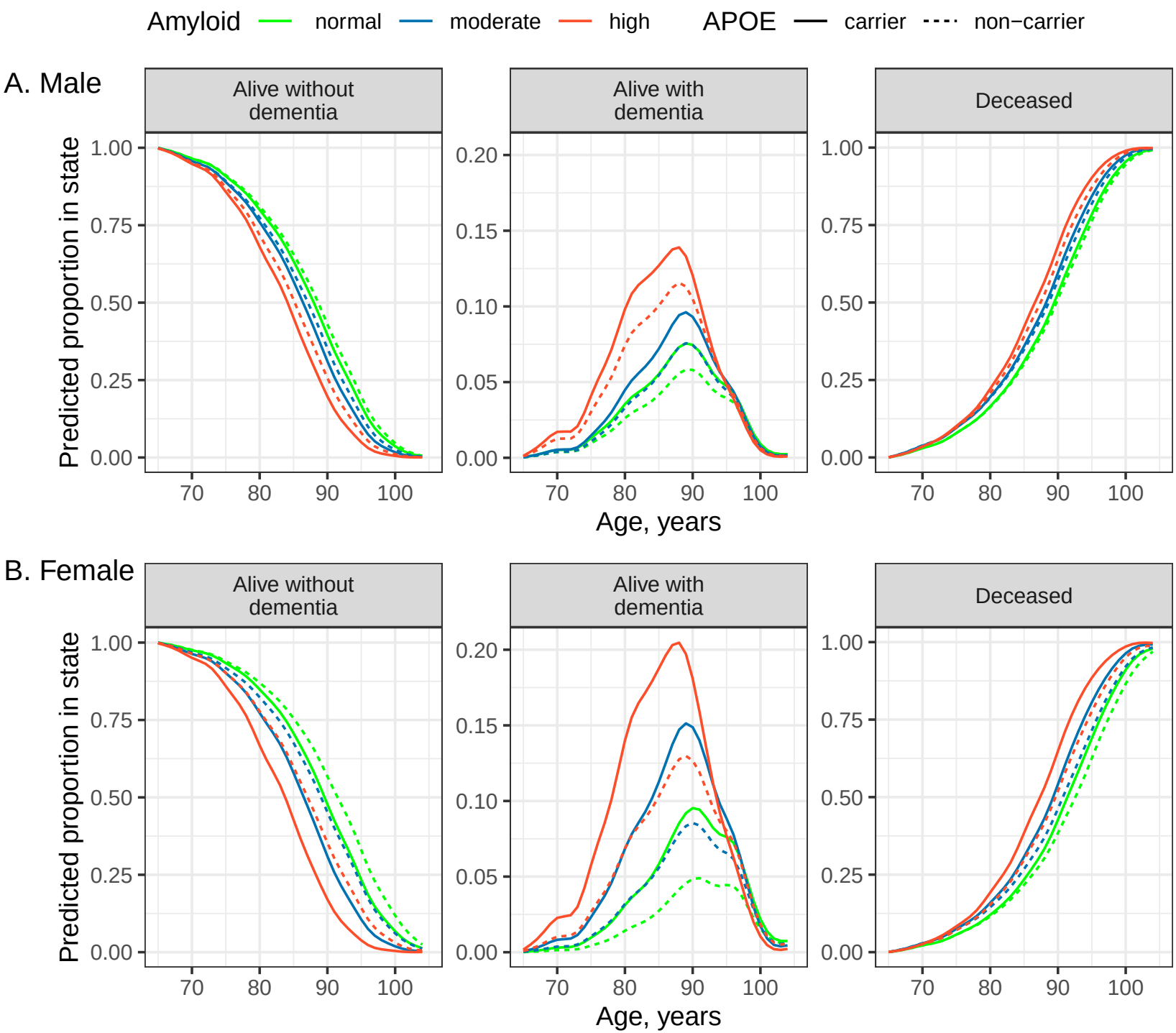
- Population based study in Olmsted County, Minn
- Age and sex stratified random sample
- Scheduled visits every 15 months
- Active set of approx 3000
- Embedded in the Rochester Epidemiology Project
  - Started in the early 1970s
  - Record linkage involving all providers of care in Olmsted County

Dementia sub-study

- Time to dementia and death, subset to age 60+
- 5080 subjects, 713 dementia, 1935 deaths
  - Over 1/2 of the endpoints occur after the cessation of active follow-up
- Primary goal is to understand the diagnostic importance of amyloid level.
- Covariates of amyloid burden, APOE, sex, education, CMC


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)


Dementia
0.0
0.1
0.2
0.3
0.4
0.5
65
70
75
80
85
90
95
100
Age
F−
F+
M−
M+


Amyloid
normal
moderate
high
APOE
carrier
non−carrier
A. Male
Alive without dementia
Alive with dementia
Deceased
Predicted proportion in state
1.00
0.75
0.50
0.25
0.00
0.20
0.15
0.10
0.05
70
80
90
100
Age, years
B. Female
Alive without dementia
Alive with dementia
Deceased
Predicted proportion in state
Age, years

Results

- Fascinating sex/APOE/amyloid story (men *are* different)
- Death
  - Male death rate without dementia is 1.4 that of females
  - Male death rate with dementia is 1.3 that of females

Results

- Fascinating sex/APOE/amyloid story (men *are* different)
- Death
  - Male death rate without dementia is 1.4 that of females
  - Male death rate with dementia is 1.3 that of females
- This interplays with dementia: separate sex effects on hazard ratios, lifetime risk, probability in state and time in state

Methods

- Counting process data set
- Multi-state hazards model
- Age as the time scale
- R survival package (3.2-9)

Counting process data

- Each subject can have multiple rows
- Every row has an id, time interval, covariates, current state, and the transition (if any) at the end of the interval
- Key rule: every person describes a physically possible path:
  - No overlaps, e.g. (64, 75] (72, 81] (81, 84]
  - No gaps, e.g. (67, 69], (80, 83]
  - No zero length intervals
  - Consistent states

Valid hazard models

- Interactions? (additive)
- Linear?
- Proportional hazards?
- Informative censoring?
- Immortal time bias
  - Cannot peek into the future
  - Covariates ("ever demented")
  - Inclusion (only those with at least 1 transition)
  - Outcome (two codes 90+ days apart)

Absolute risk

- Our sole focus on the HR is a bad idea
- With 1 arrow I can predict the absolute risk
  With 2 arrows, I can sort of guess
  With >2 arrows, I have to draw the absolute risk curves
- HR hints at underlying biology, absolute risk $=$ consequences of the biology

Absolute risk

- Our sole focus on the HR is a bad idea
- With 1 arrow I can predict the absolute risk
  With 2 arrows, I can sort of guess
  With >2 arrows, I have to draw the absolute risk curves
- HR hints at underlying biology, absolute risk $=$ consequences of the biology
- Causal
  - Predictions can be evaluated: Pr(event in 3 years)$=$ 24%
  - Group estimate $=$ average(per subject estimate)

Code

```
Aalen-Johansen
  survfit(Surv(age1, age2, event) ~ apoepos + male,
                    id = ptnum, data = data2)

Multi-state hazards model
  coxph(list( Surv(age1, age2, event) ~ apoepos + male + icmc,
                  1:2 ~ apoepos * male),
                  id = ptnum, data=data2)

 Absolute risk
   dummy <- data.frame(apoepos = c(0,0,1,1), male = c(0, 1, 0,
                                      icmc= 2)
   curves <- survfit(coxfit, newdata = dummy)
   plot(curves, ...)

 Checks
   survcheck( Surv(age1, age2, event) ~ 1, id = ptnum,
                data = data2)
```

Open issues

- Coefficient explosion
- Time dependent covariates + absolute risk
- Interval censoring and irregular measurements
- Random effects

- Simple models +
- Simple tools
- Goes surprisingly far