Introductions

Terry Therneau

25 Sept 2024

1 Introduction

- Mayo Clinic is a teriary care center, so the majority of medical questions that have come to me are "how long until": waiting time on the liver transplant list, mild cognitive impairment (MCI) and dementia, duration of physical therapy rehabilitation after a stroke, accumulation of metabolic comorbidities, etc.
- I work in support of medical research, only.
- Every feature in the survival package is a response to a research question that I have faced.
- My goal here is to trade ideas and to learn.
- Class materials on gihub.com/therneau/leidenworkshop.
- History
  - 1975, BA St. Olaf College
  - 1976-79 programmer
  - 1979-83 PhD, Stanford
  - 1983-85 Asst Professor
  - 1985-24 Mayo Clinic
- Languages
  - main: Fortran, Basic, Focal, APL, PL/I, C, awk, lex, yacc, assembler*4
  - stat: BMDP, SAS S, Splus, R, (minitab, SPSS, matlab)
  - OS: IBM 11/30, RSTS, TOPS20, JCL, Wylbur, CMS, Unix (Bell, Berkeley, Sun, Linux), Windows
  - code: Panvalet, SCCS, rcs, cvs, svn, mercurial, git

2 Models

- at each event time there is a drawing for the winner
- each obs has $r_i = \exp(\eta)$ tickets
- P(subject i wins) $= r_i / \sum_{atrisk} r_j$

The three most popular models in statistics are

- Linear: $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$
- GLM: $E(y) = g(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots)$
- Cox: $\lambda(t) = g(\beta_0(t) + \beta_1 x_1 + \beta_2 x_2 + \ldots)$
- Why? Simplicity.
  - If 'x1= weight', then β_1 is *THE* effect of weight, independent of any other variables in the model.
  - Statisticians like this.
  - Investigators really like this (a single p-value)
- Generalized additive models will replace one of the βx terms with $s(x)$, but retain the separability.

I argue that successful statistical models have 4 attributes

1. Simplicity: in the sense described above, leading to simple explanations for the effect of key predictors.
2. Statistical validity: the model must describe the data adequately. "All models are wrong. The practical question is whether a model is wrong enough to not be useful." George Box
3. Numerical stability: the code to fit a model does not require hand-holding or fiddling with tuning parameters: it just runs.
4. Speed

The transform g gets chosen to fit criteria 3; if it helps with criteria 2 that is mostly luck. (It nearly always impedes interpretability). "The reason that casinos use odds is that no one really understands them." The Cox model uses $g(\eta) = \exp(\eta)$ simply because it prevents negative values (the dead coming back to life). However, to my great surprise, the biology of this assumption often works out as well, see figures 1 and 3. Do *not* take this for granted.

Figure 4 shows that the year effect is not as simple.

```
> hip2 <- with(hips, data.frame(year=rep(year,2), age= rep(age, 2),
                                pop= c(pop.f, pop.m),
                                count= c(event.f, event.m),
                                male= rep(0:1, each= nrow(hips))))
```

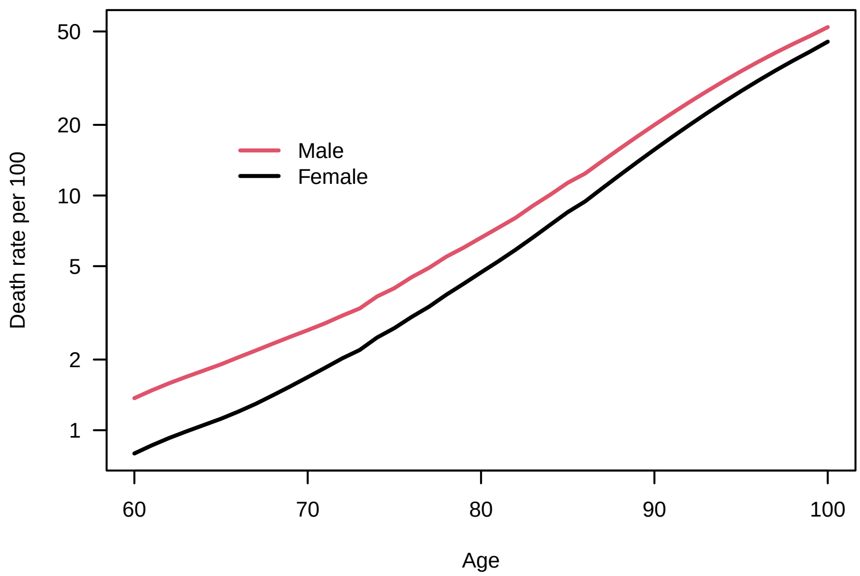

Figure 1: United States death rates by age and sex, 2020.

```
> gfit <- glm(count ~ year + age + offset(log(pop)), family= poisson,
             data=hip2, subset= (age >39 & age <101 & year>1949 & pop>0))
> summary(gfit)

Call:
glm(formula = count ~ year + age + offset(log(pop)), family = poisson,
    data = hip2, subset = (age > 39 & age < 101 & year > 1949 &
        pop > 0))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.950989   3.759939  -1.317   0.1879
year        -0.004397   0.001916  -2.295   0.0217
age          0.112989   0.001905  59.321   <2e-16

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 8332.9  on 5076  degrees of freedom
Residual deviance: 3752.3  on 5074  degrees of freedom
AIC: 6709.3

Number of Fisher Scoring iterations: 5
```

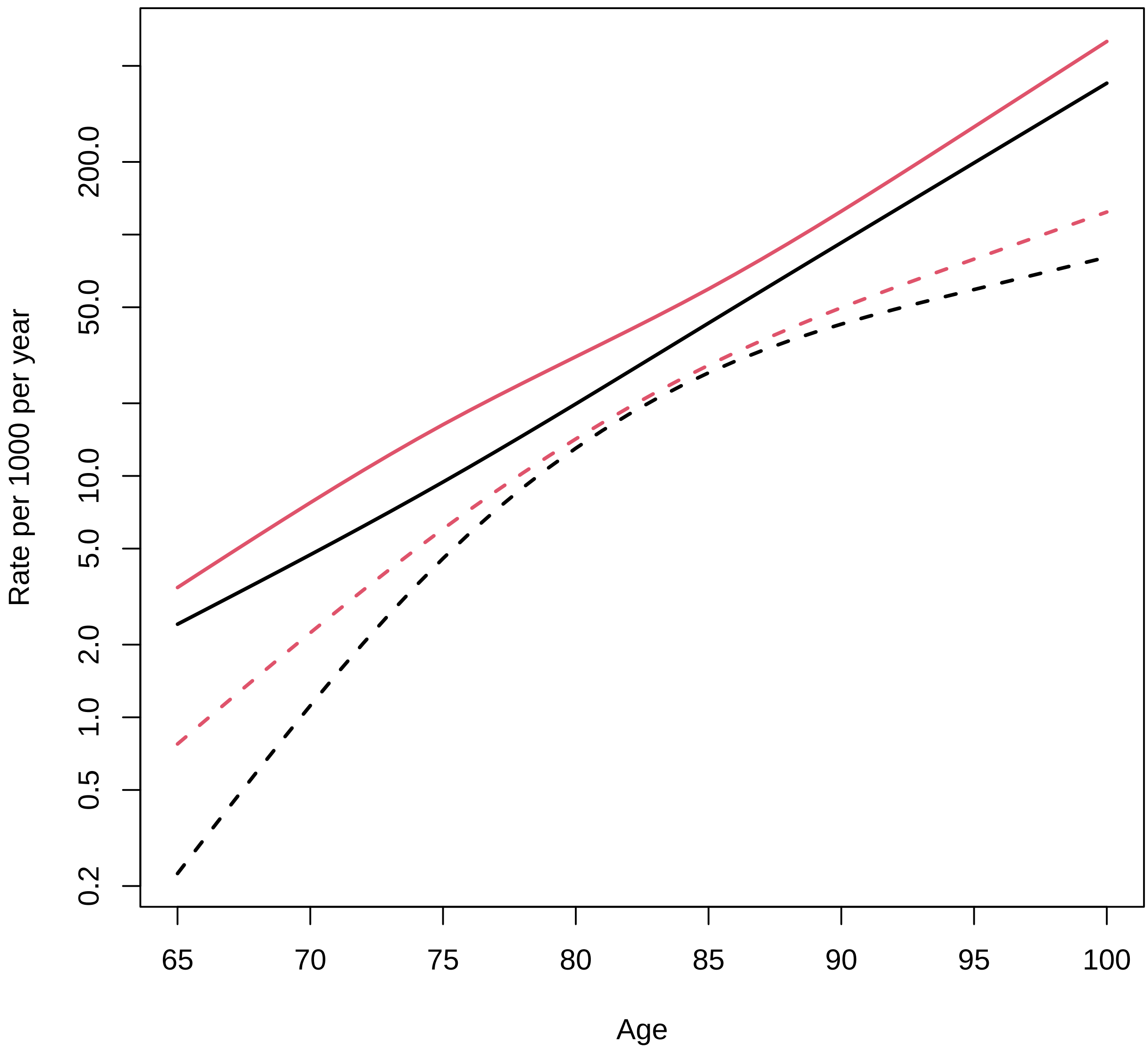

Figure 2: Death and dementia rates from the Mayo Clinic Study if Aging.

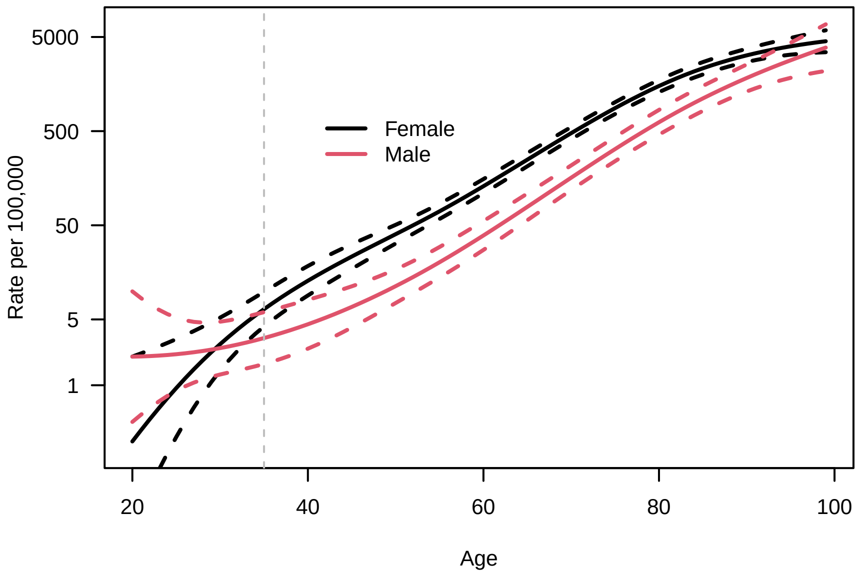

Figure 3: Hip fracture rates in Olmsted County, Minnesota, 1929-1992.

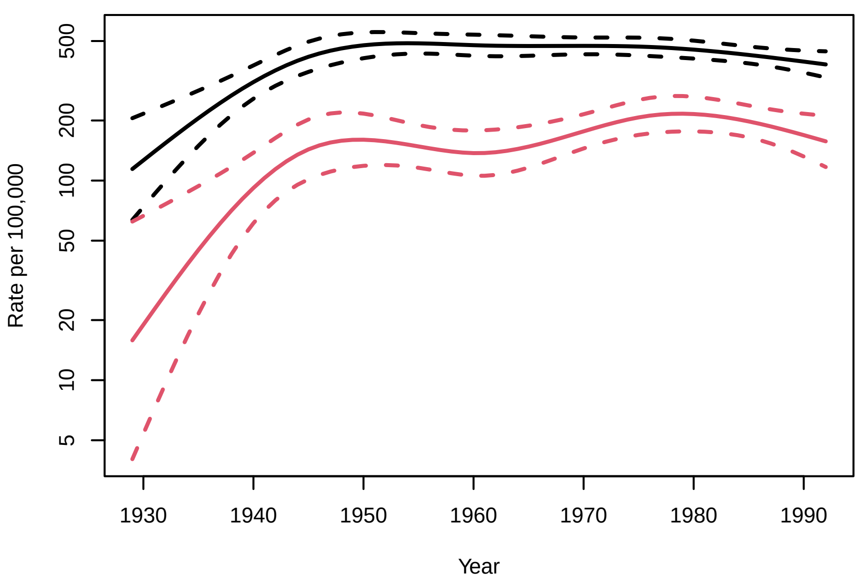

Figure 4: Hip fracture rates in Olmsted County, Minnesota by year.