An accidental statistician

Terry Therneau

Oct 2024

History

- 3/1953: born Preston Minn. Eldest of 9, grew up on a dairy farm
- 71-75: St Olaf College, BA Mathematics
- 76-79: Mayo Clinic, programmer
- 79-83: Stanford, PhD
- 83-85: University of Rochster, Asst Prof
- 85-xx: Mayo Clinic
- Family:
  - Adam 9/78, Natan 9/81, Joel 9/83, Isaac 5/85, Elizabeth 8/87
  - Cora, Cossette, Remi, Effie, Sylvan, Maraia, Clarissa

Mayo

- My job is to support medical research
  - No teaching
  - No graduate students
- Biostatistics is staffed approximately 1:2:3 with PhD, MS, BA statistical personnel.

Success (how?)

- I found that I was rewriting "The seven habits of highly effective people"
  - Where are you going (be proactive, begin with the end in mind, first things first)
  - Interdependence (win/win, first understand, teamwork)
  - Sharpen the saw
- He speaks and writes better than I do
- Messages are not new

Really great moral teachers never do introduce new moralities: it is quacks and cranks who do that. . . . The real job of every moral teacher is to keep on bringing us back, time after time, to the old simple principles which we are all so anxious not to see; like bringing a horse back and back to the fence it has refused to jump or bringing a child back and back to the bit in its lesson that it wants to shirk. C. S. Lewis

What I seem to have done right

- Luck
- Focus on the goal (medical research)
  - Take time to think
  - Research as needed
- Work with good investigators
- Work with a good statistics team
- Outlook not installed on my phone

Difficut

- Instatiable demand
- Need to say “no”
- The tyranny of the urgent
  - abstract deadlines
  - grant deadlines
  - review deadlines
- Administrivia

Farming lessons

- You are never *ever* going to get everything done
- Pick a job, and finish it
- Take time off

- Lila Elveback. “I have seen one Gaussian distribution in my work: the height of adult males.” (Health, Normality, and the Ghost of Gauss)
- Ken Offord. “A B+ is no longer good enough. These are real patients, the answer has to be right.”

Residuals

- Continuous ejection fraction (Moss) vs categorical (Bigger)
- "If this were a linear model, I'd use residuals"
- Led to martingale residulas paper 3 years later
- 20 years later: M-resids are useful, but not for that

Multi-Center Diltazaem Post Infarction Trial (MDPIT)

- Calcium channel blocker to prevent second events
- Sample size of 243 events
- Rough cost of $25,000 per event
- Some people have 2, can we make use of the second event?
- 5 years later, robust variance for the AJ
- still a key part of the survival package

Survival package

- Written in reaction to needs (residuals)
- First statlib release in 1989. "Might be useful. . . "
- Why the success? + First
  - Every feature in the package is the reaction to a dataset
  - We use it every day
  - I work to make the results provably correct (Don't do everything, but do it right)
  - Efficiency
  - coxph(Surv(time, time2, status) ~ x1 + x2)
  - original: $O(n * d) + O(dp + dp^2)$ = bookkeeping + statistics
  - now: $O(2np + dp^2)

Multiple endpoints

- UDCA trial in Primary Biliary Cirrhosis, PI Keith Lindor
- Very slow progression. After 3+ 5 years, 10 vs 6 deaths.
- Look at liver outcomes: death, liver transplant, histologic progression, esophogeal varices, encephalopathy, ascites, (doubling of bilirubin).
- 46 have 1, 14 2, and 3 have 3 endpoints. More information?
- Andersen-Gill vs Wei-Lin-Weissfeld vs Prentice-Williams-Petersen models.

Random effects Cox models

- Original study conducted by Elving Anderson, U of Minnesota.
- From 1944 to 1952, 544 sequential breast cancer cases seen at the University Hospital were enrolled, with information gathered on parents, sibling, uncles, aunts, and grandparents.
- In 1991 the study was revived by Dr Tom Sellers. Updated pedgrees were obtained for 426 (13351 males, 12699 females)
- Genesis of the kinship and coxme packages; variance of $\sigma^2 I + \tau^2 K$
- (Assume a data set with id, father id, mother id, family id, sex. Can you automatically draw a pedigree for family i?)
- Later use in monitoring of liver transplant centers.

Mean restricted time in state

- A study with Dr. William Brown in the stroke center.
- Question: what factors influence the time to completion for a stroke rehabilitation program?
- Fit the usual Cox model with HR, but then thought "what about mean time in state".
- He was delighted. It has a direct interpretation wrt reimbursment.
- Outcome: simpler interface + confidence intervals

Analysis of FTICR mass spectrometry data

- Work with Ann Oberg, PI David Muddiman
- Very different data, large learning curve
- After finally "getting our arms wrapped around" the problem, the statistical answer was — ANOVA.

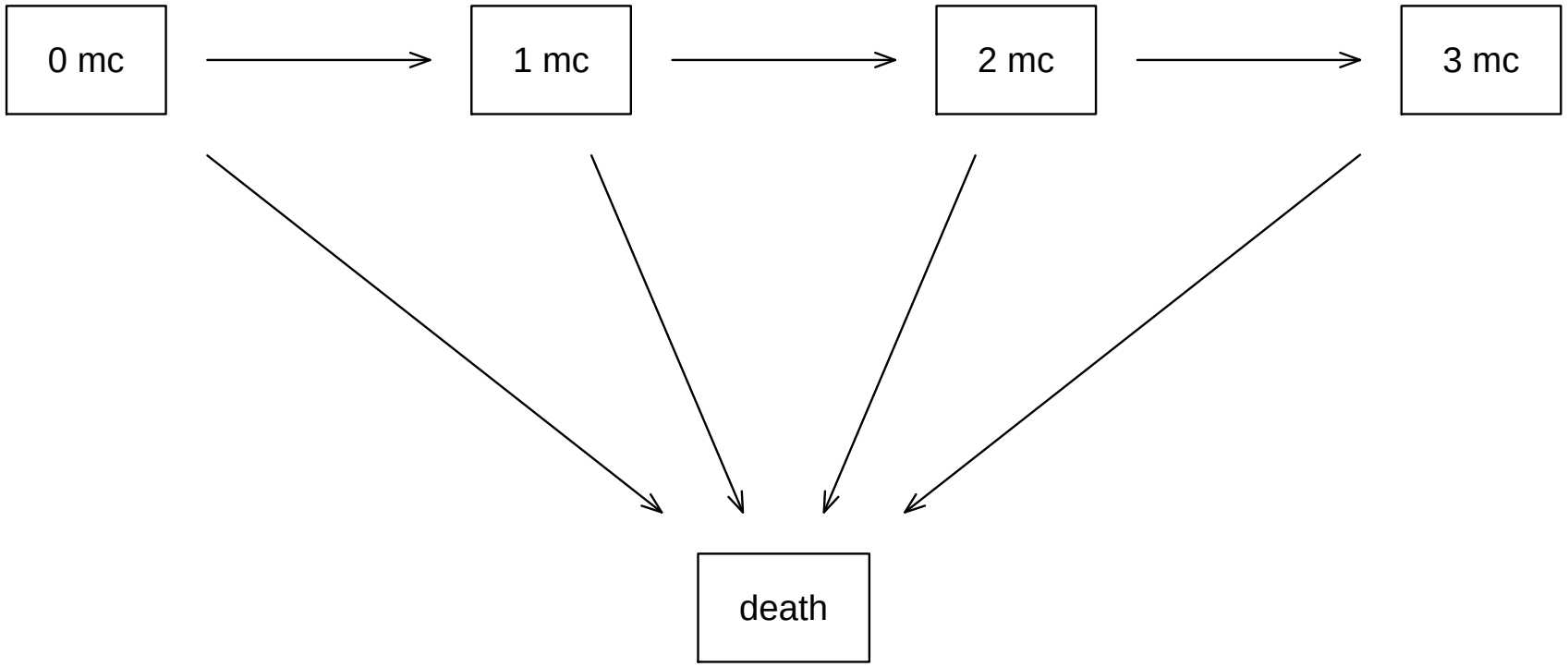
Evolution of coxph

```
coxph(Surv(time, status) ~ x1 + x2 + strata(inst), mydata)
coxph(Surv(time1, time2, status) ~ x1 + x2, mydata)
coxph(Surv(time1, time2, state) ~ x1 + x2, id=clinic, mydat
coxph(list(age1, age2, state) ~ male + apoepos + cmc + amyl
       1:2 ~ male *(apoepos + amyloid)), mydata, id= clinic

nfit <- coxph(list(Surv(age1, age2, state) ~ nafld,
                1:5 + 2:5 + 3:5 ~ male / common),
                data=ndata, id= id, istate= cstate)
```

Multistate hazard models

- Fatty liver disease research with Dr. Alina Allen.
- All diagnosis of MAFLD, with 4 age/sex/year matched controls
- Progression of metabolic comorbidities: diabetes, hypertension, hyperlipidemia
- Analysis on age scale




```
nfit <- coxph(list(Surv(age1, age2, state) ~ nafld,
                   1:5 + 2:5 + 3:5 ~ male / common),
                   data=ndata, id= id, istate= cstate)
```

Mayo Clinc Study of Aging

- age and sex stratified random sample from Olmsted County, Minn
- strata of 50-59, 60-69, 70-79, 80-89
- oversample older ages
  - multistate hazard models,
  - multinomial models
  - hidden Markov models
  - non-linear mixed effect (GLME) models, ...
- understanding 30 year processes with 3-5 years per person