Multiple Event Models
Some Practical Aspects

Terry Therneau
Divison of Biostatistics
Mayo Foundation

March 2005

- Mixed up models
  - multiple states
  - multiple visits/state
  - A clear win
- Directional models
  - multiple states/patient
  - no returns
  - Very useful for summaries
  - Perhaps for modeling
- Repeated events
  - one type of outcome
  - deterministic path
  - Savings in sample size
  - . . . but worth it?

Crohn's disease

- a recurrent inflammatory disease of the gut
- of uncertain origin, but possibly immune-related
- frequently intestinal obstruction and abscess formation
- a high rate of recurrence

- treatment
  - anti-inflammatory (steroids, immunosuppression)
  - surgical removal of inflamed sections

The study

- 174 Olmsted County, Minnesota, residents, 1970–1993.
- 2 months to 24 years of follow-up ($\mu = 7.3$)

State	Num Ever	Entries to State					
		1	2	3	4–5	6–9	10+
Remission	146	40	49	28	16	10	3
Mild	138	50	32	20	23	11	2
Severe							
Drug-resp	64	44	12	2	4	2	0
Drug-dep	42	29	5	6	2	0	0
Drug-ref	45	29	13	1	2	0	0
Surgery	100	64	20	7	7	1	1
Postsurgery rem	85	55	22	6	1	0	1

Analysis

- Creating an analysis data set is fairly simple
- Cox models are interpretable in terms of rates of transition
- Questions
  - Do the gender's differ?
  - Is the model Markov?
  - What are the recurrence times, and durations in state?
  - ...

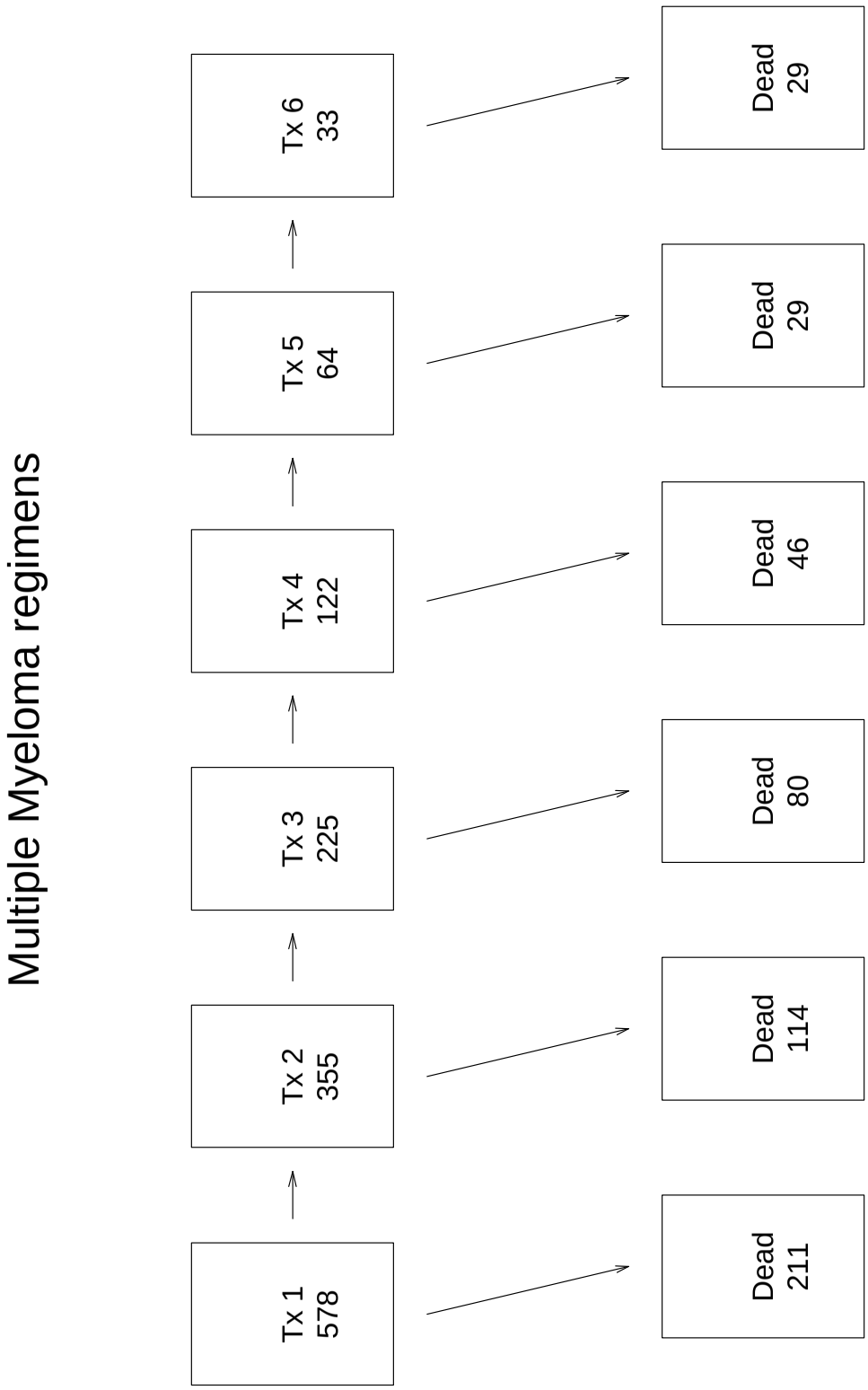
Conclusions

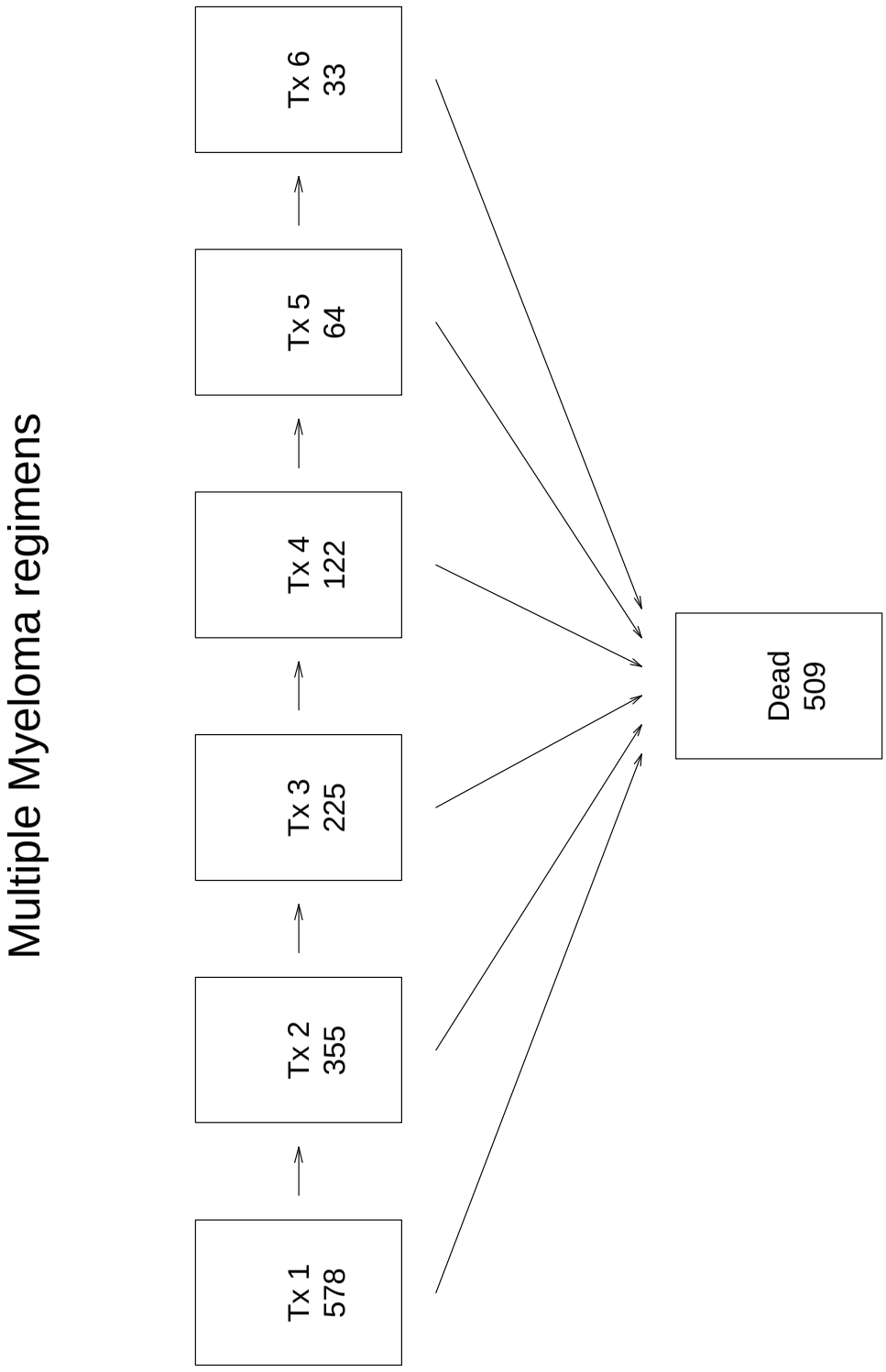
- One *has* to use a multi-state model of some sort
  - Cannot follow unique paths (too many)
  - Sample size demads reuse
- Markov transition model assumes exponential waiting times (constant hazard) — untenable
- The multi-state Cox model works very well
- Data sets like this are rare as hen's teeth

Directional models

Multiple possible paths for a patient.

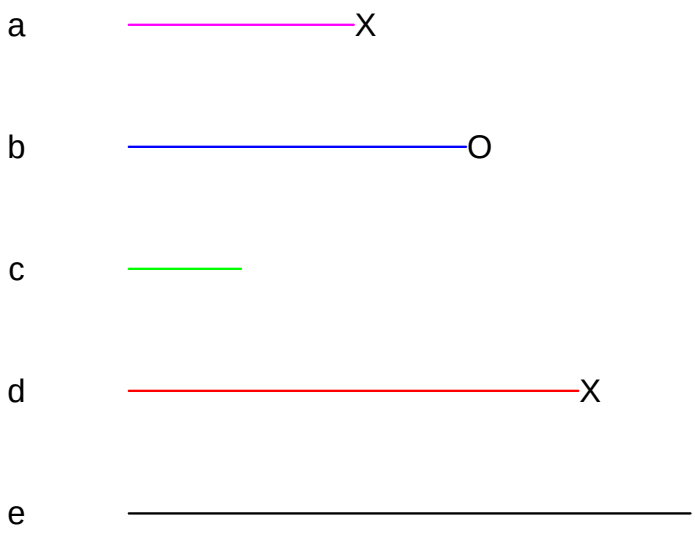
No returns to a previous state.

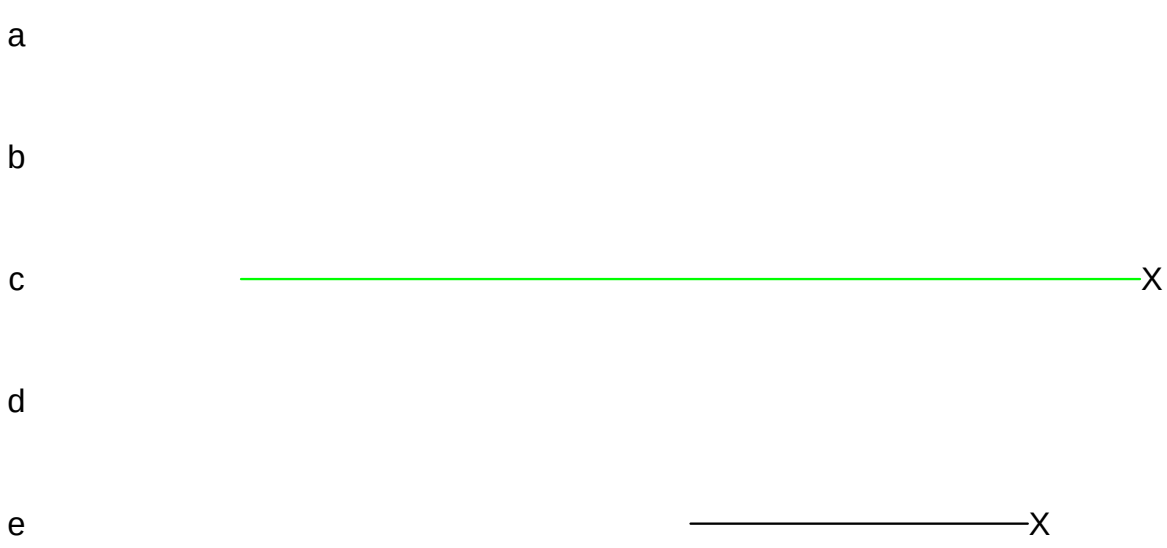

Multiple Myeloma regimens
Tx 1
578
Tx 2
355
Tx 3
225
Tx 4
122
Tx 5
64
Tx 6
33
Dead
211
Dead
114
Dead
80
Dead
46
Dead
29
Dead
29


Multiple Myeloma regimens
Tx 1
578
Tx 2
355
Tx 3
225
Tx 4
122
Tx 5
64
Tx 6
33
Dead
509

Question: What fraction of patients reach each state?

- No censoring: 578/578, 355/578, 225/578, . . . 33/578
- With censoring: redistribute-to-the-right algorithm.
  - At the time of censoring, redistribute the weight to an appropriate cohort.
- This solution is related to the *cumulative incidence* function
  - Survival curve: number who will reach an endpoint, assuming all other endpoints are removed
  - CI curve: number who will actually reach the endpoint


a
X
b
O
c
d
X
e

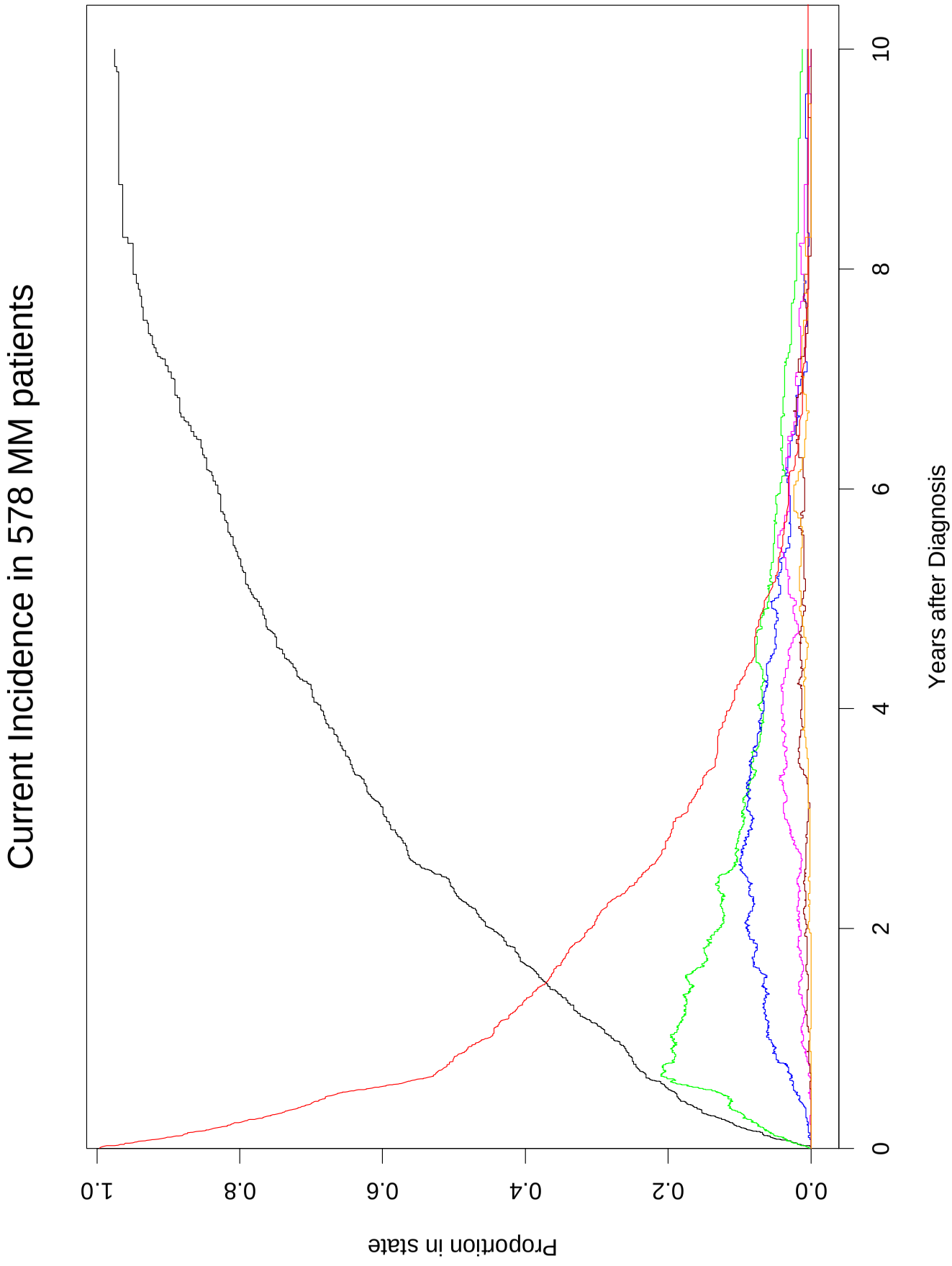

a
b
c
X
d
e
X

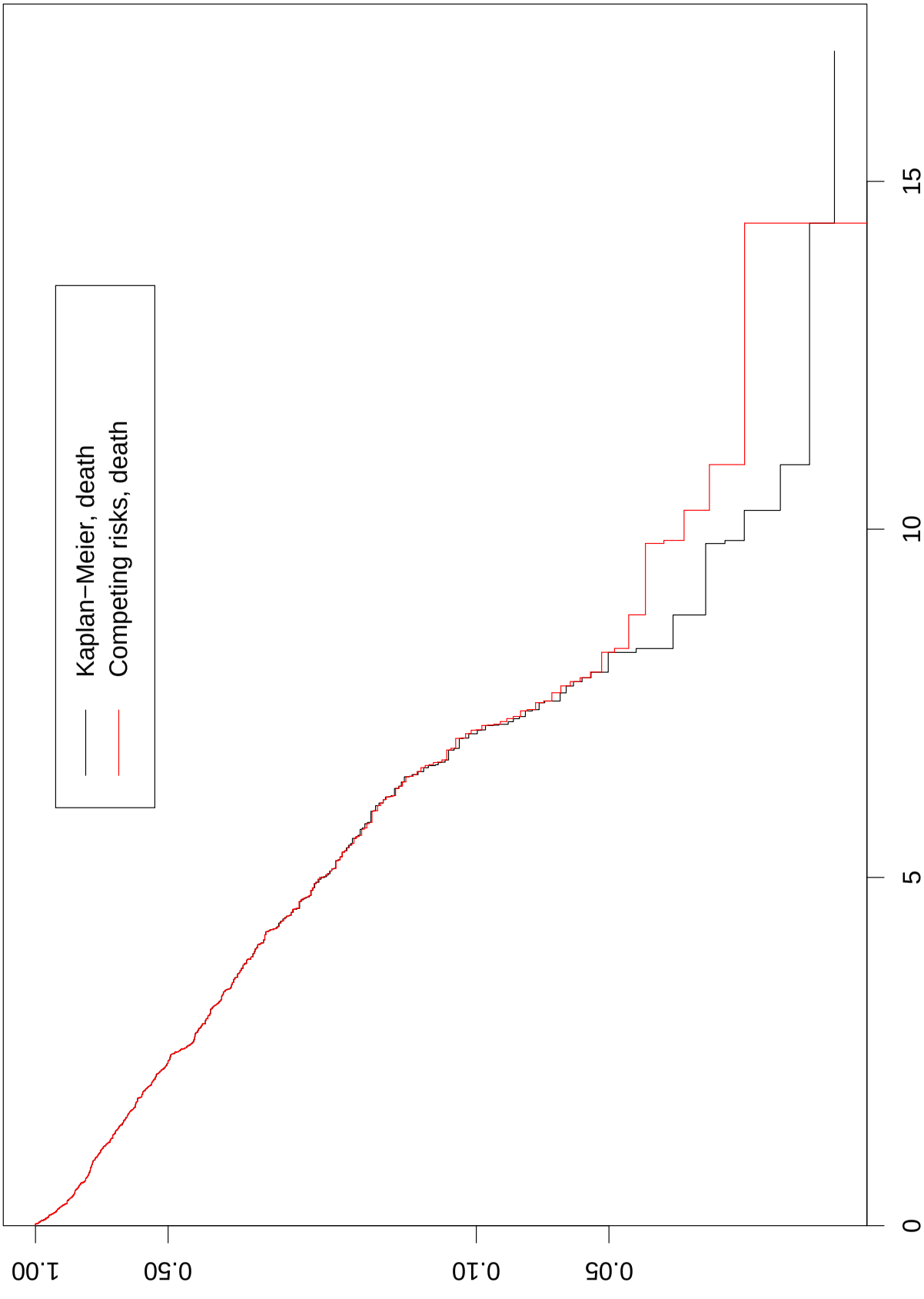
Current incidence function

If there were no censoring, this would just be the number in each state, at each point in time.

```
fit <- survfit.ci(Surv(xtime, status) ~ strata(state),
                  id= clinic, data=mm1)
plot(fit, col=1:6, mark.time=F, fun='expect')
```

- *status* =1 if a transition occured at time *xtime*
- *state* = number or name of ending state (value ignored if censored)
- *id* = subject id, for multi-state transitions


Current Incidence in 578 MM patients
Proportion in state
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Years after Diagnosis

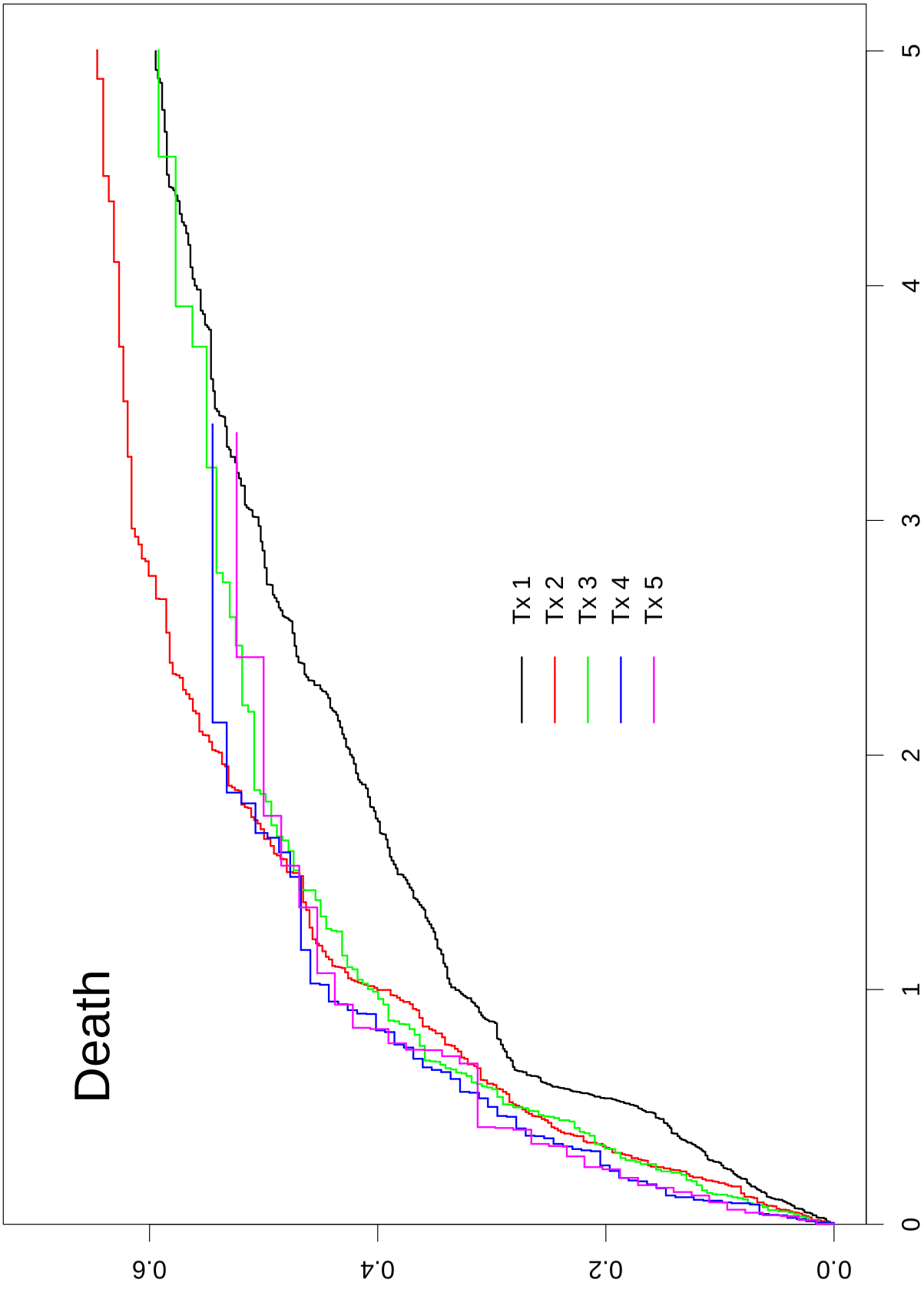

Kaplan−Meier, death
Competing risks, death
0.05
0.10
0.50
1.00
0
5
10
15

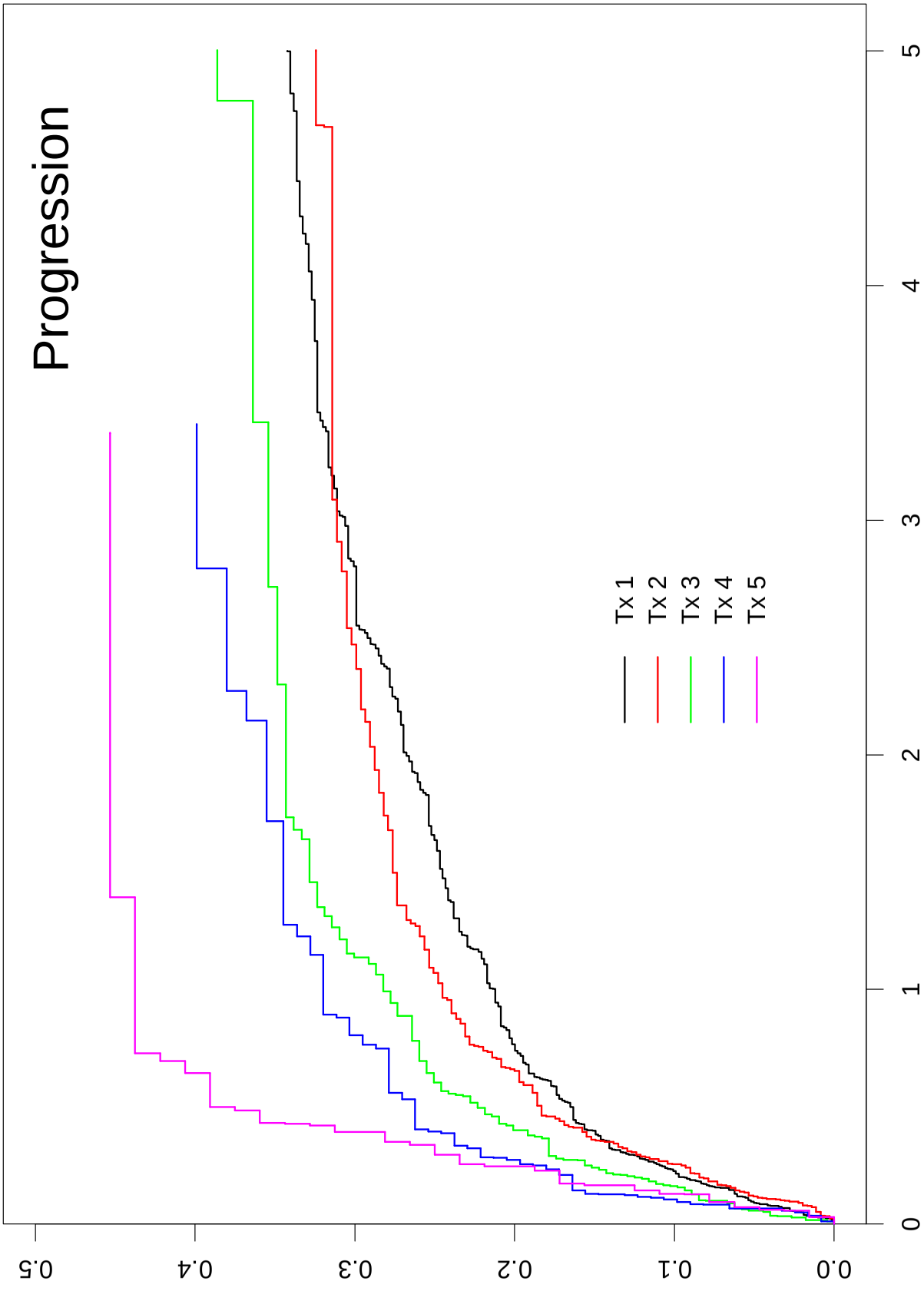
The number of people who get to state 5–6 is small

So small that the plot is not readable.

It is useful to rescale:

- Competing risks for death or transition, Tx 1
- Competing risks for death or transition, Tx 2
- ...
- Competing risks for death or transition, Tx 5


Death
Tx 1
Tx 2
Tx 3
Tx 4
Tx 5
0.0
0.2
0.4
0.6
0
1
2
3
4
5


Progression
Tx 1
Tx 2
Tx 3
Tx 4
Tx 5
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5

Cumulative Incidence Plots

What proportion of patients *will actually appear at my office* in any given state?

	18 months after initiation		
	% Dead	% Failed	%On Rx
Tx 1	40	25	35
Tx 2	48	28	24
Tx 3	47	33	20
Tx 4	48	34	18
Tx 5	47	45	08

At 18 months

- Dead: 4/10, 1/2, 1/2, 1/2, 1/2
- Prog: 1/4, 1/4, 1/3, 1/3, 1/2

Models

- The components of CI curves are hazards and survival
  - Available from Cox models
- Relative risk: use standard models
- n can get really small
- Which coefficients?
  - Is the effect of `x=age` the same for Tx 1 $\rightarrow$ Death and Tx 2 $\rightarrow$ Death?

Directional models: Conclusions

- Very useful for summary overviews
- Models may be weak
  - Small n
  - State by coefficient interactions
  - Subject selection (unmeasured covariates)
- Data sets are uncommon.

The rest of the data

Common sizes

	N	0–1 event	2 events	3 events	4+ events
Bladder cancer	86	57	7	8	14
Cystic fibrosis	645	564	53	20	8
CGD	128	111	9	5	3
Cardiac	2466	2428	38	-	-

"Multiple event" data sets often have <15% with multiples!

How much juice *can* you squeeze out of a lemon?

The good

In many data sets, each extra observation is worth 1/4 to 1/3 of a "new" case.

It is easy to do the analysis.

Choice of setup

Anderson-Gill style

- "Mixed-up" model
- Advantages
  - data does not run out
  - stable coefficients
  - little bias due to selection
- Disadvantages
  - Is it really true?

Conditional model (PWP)

- Directional model
- Advantages
  - For acute disease, states are not the same
  - Easy to investigate state*covariate interactions
- Disadvantages
  - run out of data *very* fast
  - major selection biases

Not so good

Consider the CGD study.

- 128 subjects, 44 first events, 32 "extra" events.
- A potential gain of 73% more information.
- With all the right covariates:
  - each extra event is worth about $\sim 25\%$ of a new one
  - treatment se drops from .34 to .31
- Nearly ideal study
- Was it worth it?

Bladder cancer study

- 86 subject, 47 first events, 65 "extra" events.
- Treatment se reduced from .30 to .28

Old joke:

"A conservative is someone who believes that nothing should ever be done for the first time."

- Many pharmaceutical statisticians
- appear to believe
- that the FDA believes
- that no analysis should ever be done for the first time.

Summary

1. If you've got the data, use it.
2. When some states have small n, several modeling issues become problematic. Summary and display may still be useful.
3. For acute disease data, use caution.
   - but still — even if the gain is small, it's free
   - (in MDPIT, each original case cost over $50,000)