

# Inference for Multistate Data under Complex Censoring Structures

A discussion by

**Somnath Datta**

Department of Bioinformatics & Biostatistics

University of Louisville

[www.louisville.edu/~s0datt03](http://www.louisville.edu/~s0datt03)

@ JSM 2007, Salt Lake City, 7/29/07

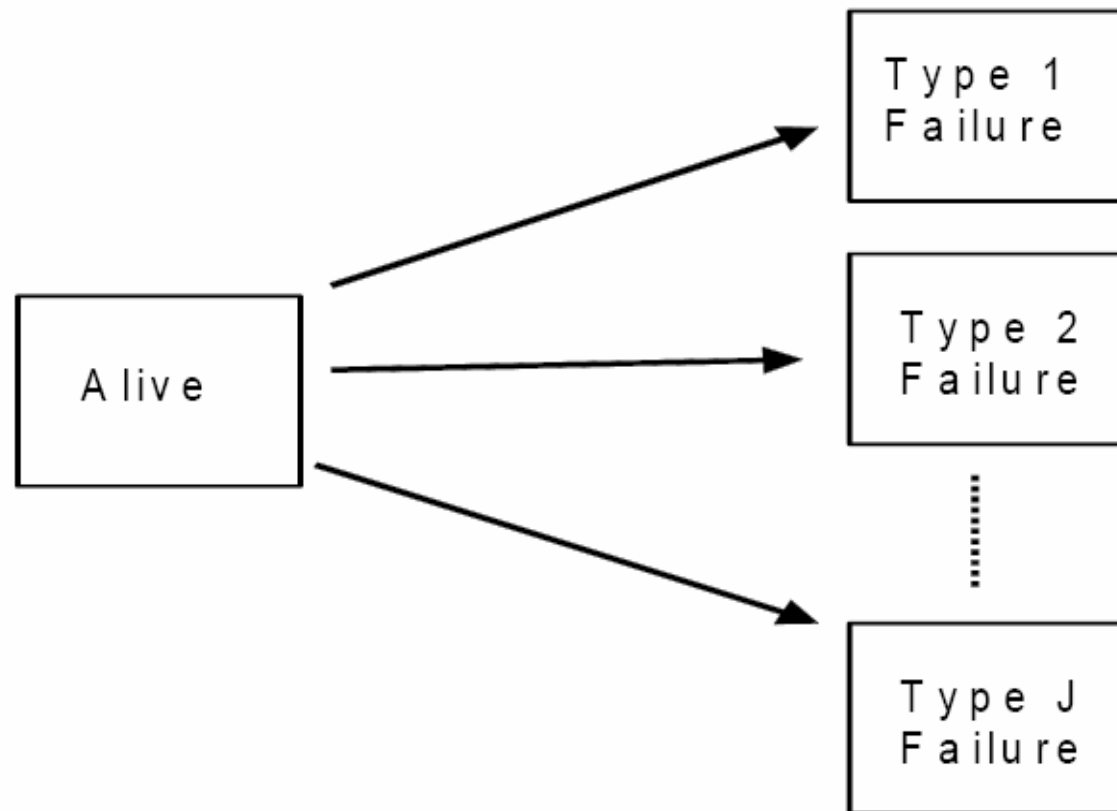
- Multistate models are generalizations of univariate survival data
- Can be represented graphically where the nodes are the states and the allowed transitions between states are the directed edges
- Progressive multistate models are acyclic and has a tree structure

# Examples

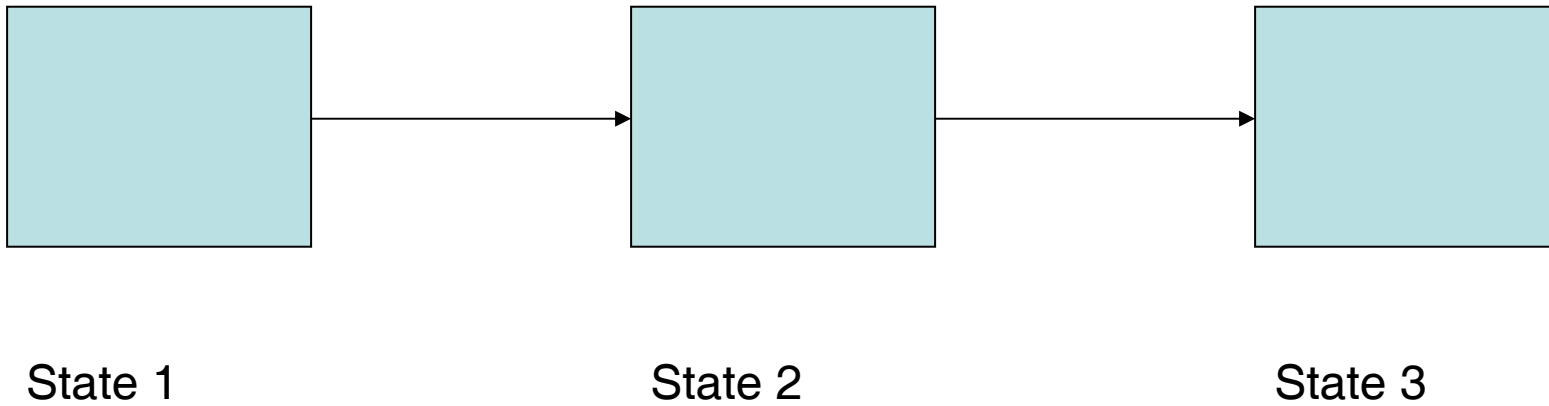
- (Univariate) Survival data



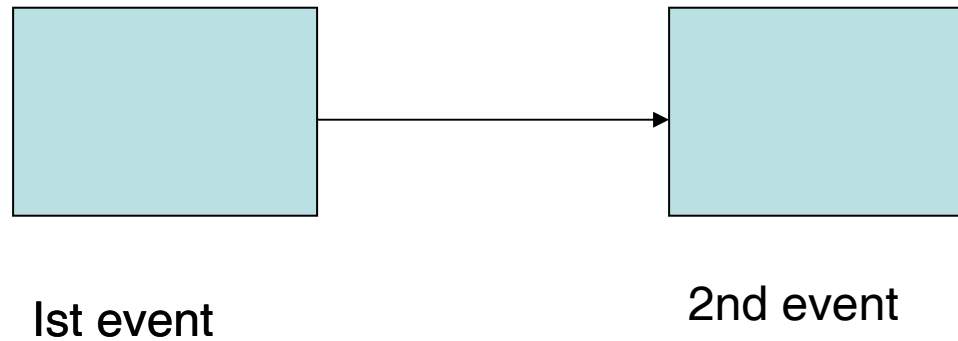
- Competing risk data (multiple decrement models)



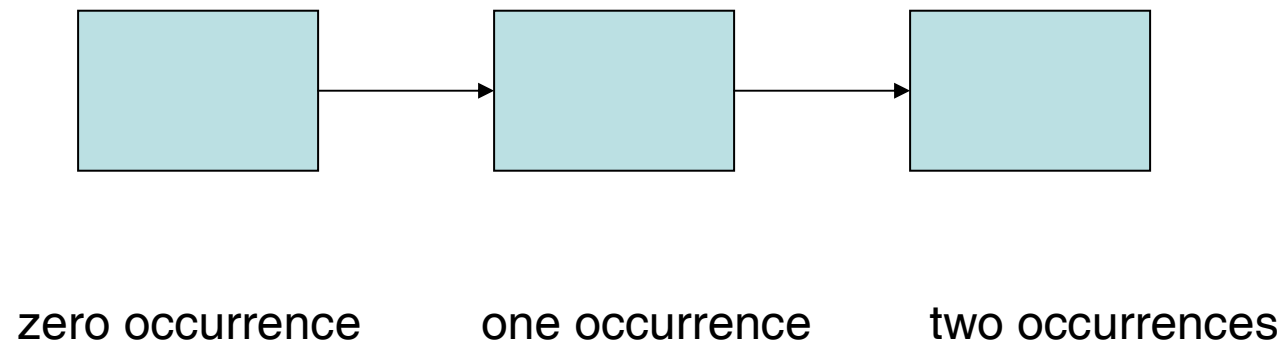
- Tracking models



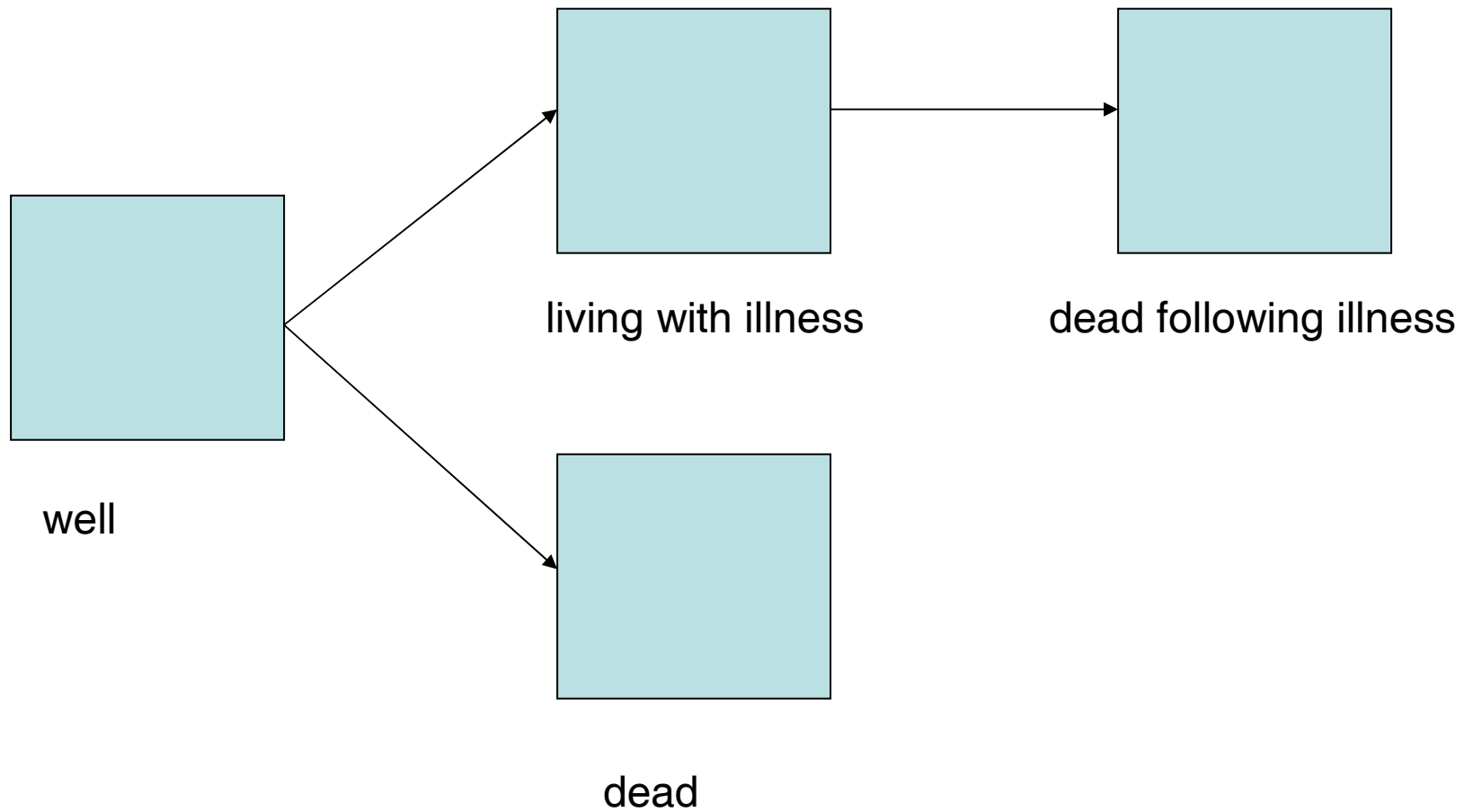
- Bivariate survival data



- Recurrent events data



- Irreversible illness-death models



# Censoring schemes

- Right censoring

individuals are followed up to an individual specific time  $C_i$

all transition times and states visited up to that time are recorded

- Current status

each individual is inspected at a single time point and the state information is recorded

- Interval censoring

Each individual is inspected at multiple time points and the corresponding states are recorded

This (along with the model topology) leads to intervals containing each transition time

# Inference for marginal model functionals

- State occupation probabilities
- State entry and exit time distributions
- State waiting time distributions

# Inferential approaches

- Parametric

parametric distributions for transition times + structural assumptions (e.g., Markov or semi-Markov)

- Nonparametric

NPMLE: possible in some situations  
often problematic

- Nonparametric

approximate unbiasedness (e.g. smoothing)  
for (full) complete data answers

In principle, ad hoc estimators can be made  
efficient towards a model

Robins

- Higher order efficiency

Are there any limiting factors in practice?

e.g., amount of data, computational aspects,  
selection of the kernel order

Comments?

# Modeling effects of covariates

- Cox's regression
- Aalen's linear hazard
- Proportional odds
- Accelerated failure times
- Nonparametric models

Could be studied on various transition/event times, state waiting times etc

# Inferential approaches

- Parametric

(pseudo) likelihood + EM algorithm

e.g., interval censoring with fixed inspection times

- Semiparametric

partial likelihood, imputed (or integrated)

partial likelihood, other rank based approaches

- Nonparametric

For a single categorical (e.g., discrete) covariate  $Z$ , split samples by values of  $Z$ , compute mean function from estimates of cdf's in each group.

For continuous covariates, smooth the above answers.

For multiple covariates use additive models where each term is non-parametrically estimated.

# Other issues

- Left truncation (+ interval censoring)

eg, dynamic cohort in epidemiologic studies

Likelihood needs to be modified to account for left censoring

- Missing covariates

Estimate/predict the missing covariate values from observed covariates and use it in answers with known covariates.

Needs separate models + adjustments of variance estimates

Finkelstein

- Other approaches?

Model quantities (e.g., hazards or the odds)  
directly using available history at time  $t$  which  
includes covariates observed prior to time  $t$

Dynamic regression models

Comments?

# Fixed versus random inspection times

- For fixed inspection times, the analysis is essentially parametric since the dimension is finite

Sun, Finkelstein

- In some situations, the inspection times could be realizations of a renewal process which results in interval censoring of multi-state events by random intervals

# Possible approaches

- Consider a rank based estimating functions for complete data and use an averaged estimating equation over all rankings consistent with the observed interval censored data

Satten, 1996 (survival data)

- Impute the transition times from the observed intervals (needs modeling) and use an average estimating equation

Satten, Datta, Williamson, 1998  
(survival data)

Comments?