The Analysis of Doubly Interval Censored Data using Stochastic EM

Emmanuel Lesaffre [1 2] and David Dejardin [2 3]

[1]Department of Biostatistics, Erasmus MC

[2]I-BioStat, Katholieke Universiteit Leuven

[3]Global Biometric Sciences, Bristol-Myers Squibb

Österreichische Statistiktage 2011

Outline

Motivating example

Phase III Study in 2^{nd} line PH+ Chronic Myeloid Leukemia (CML) in Chronic phase

- Original objective of study:
  Non-inferiority of dose 100mg/day on Major Cytogenetic Response
- Factorial design:
  2 Doses (100 vs 140mg/day) of Dasatinib and 2 schedules (QD vs BID)
- **Here** variable of interest:
  Time from Complete Cytogenetic response (CCyR) to Major Molecular response (MMR)
- 670 subjects randomized
  Sample restricted to **213 subjects with CCyR and MMR**

Motivating Example

- Multiple measures of strength of disease in PH+ CML
  - Measurements taken in bone marrow and blood
  - Cytogenetic response: **no new diseased cells** in bone marrow
  - Molecular response: **response at gene level** in blood
  - MMR more accurate measure of disease strength
    $\Rightarrow$ MMR occurs after CCyR
- Response levels and time to achieve response (CCyR and MMR separately) were similar across two doses

Motivating Example

Questions

- Estimate time between CCyR and MMR

- Is time between CCyR and MMR similar across doses?

Motivating Example

Questions

- Estimate time between CCyR and MMR

- Is time between CCyR and MMR similar across doses?

Questions are of interest for treatment management:
When did subject reach maximum treatment benefit?

Disease sequence:
D $\Rightarrow$ HR $\Rightarrow$ CR $\Rightarrow$ MR $\Rightarrow$ LMR $\Rightarrow$ LCR $\Rightarrow$ LHR $\Rightarrow$ AB

Motivating Example

Issue

- CCYR and MMR measured every 3-6 mths
- Exact time of CCYR and MMR not known

- Classical situation in oncology, e.g. duration of tumor response
- Usually dealt with by assuming data **right censored**
  event/censoring time= end of observed interval

Motivating Example

Typical data collected

Subject	Dose	U_{li}	U_{ri}	V_{li}	V_{ri}
1	100 mg/m^2	0	3	2	9
2	100 mg/m^2	0	3	0	3
3	100 mg/m^2	0	3	3	9
4	140 mg/m^2	3	6	12	15
5	140 mg/m^2	4	5	3	6

U_{li}, U_{ri}, V_{li} and V_{ri} are in months

These data are not the real data (confidentiality reasons)

Right censored data

We observe $\min(T, C)$ with

T = time of interest

C = censoring time

Main methods:

- Kaplan-Meier (KM): Estimation of distribution
  - Product limit estimator
  - Also maximum likelihood estimator
- Cox Proportional Hazard (PH)

$$S_T(t|X) = S_T(t|X = 0)^{e^{\beta X}}$$

  - Partial likelihood estimation of β
  - Breslow estimator (MLE) for baseline survival function $S_T(t|X = 0)$

Interval censored data

We observe $[T_l, T_r]$, with $T \in [T_l, T_r]$ (closed interval)

Main methods:

- Turnbull: Distribution
  - Maximum likelihood estimator
  - Estimated using EM algorithm (self-consistency algorithm):
    - Determine regions of positive (probability) mass
    - Estimate mass
- Pan = extension of PH: Distribution with covariates
  - Multiple Imputation approach

Turnbull estimate-1

Intervals of possible mass

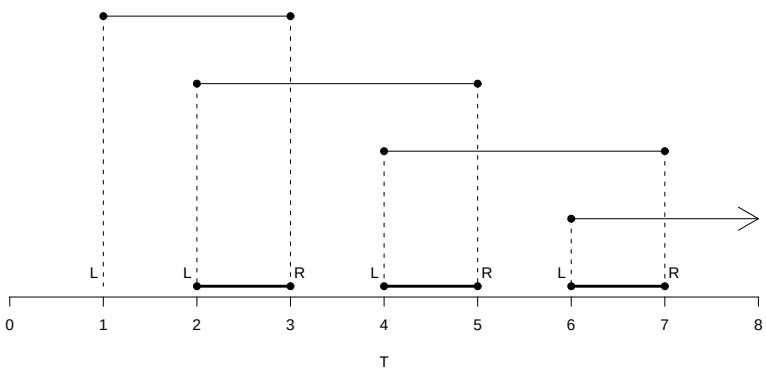
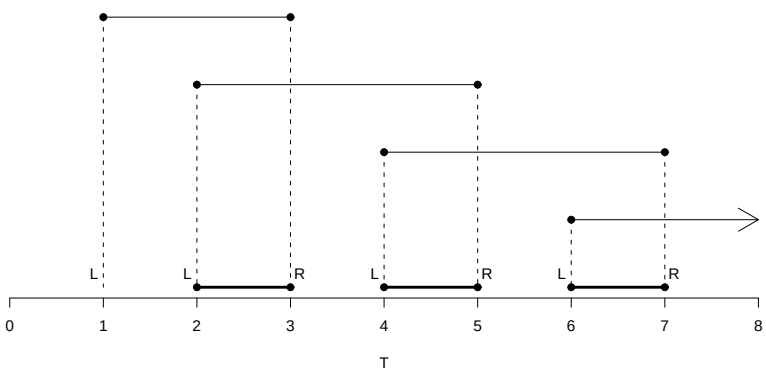


FIGURE 3.1: Determination of regions of possible support for the Turnbull estimate from observations $[1, 3]$, $[2, 5]$, $[4, 7]$ and $[6, \infty]$. The bold lines on the bottom indicate the 3 regions of possible support.

Turnbull estimate-1

Intervals of possible mass



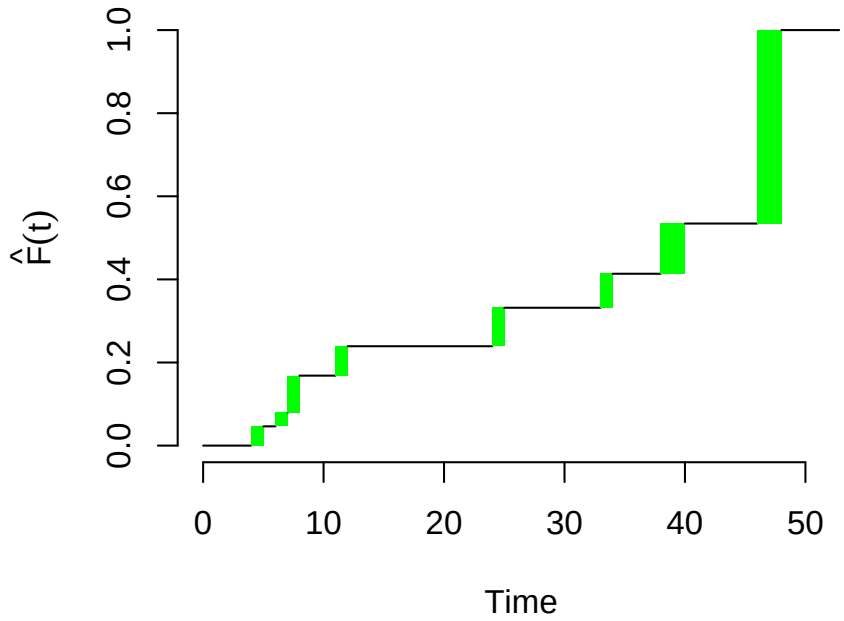


FIGURE 3.1: Determination of regions of possible support for the Turnbull estimate from observations $[1, 3]$, $[2, 5]$, $[4, 7]$ and $[6, \infty]$. The bold lines on the bottom indicate the 3 regions of possible support.

Turnbull estimate-2

Maximization part of algorithm

Once in a first step the regions of possible support are calculated, the mass assigned to each of these intervals must be estimated in a second step.

For half open or closed intervals the above reduction algorithm gives rise to a set of intervals $\{\lfloor p_j, q_j]\}_{j=1}^m$. Define $s_j = S(p_j-) - S(q_j+)$, $j = 1 \ldots, m$. Then the vector $\boldsymbol{s} = (s_1, \ldots, s_m)^T$ where $\sum_{j=1}^m s_j = 1$ and $s_j \geq 0$, $j = 1 \ldots, m$, defines equivalence classes in the space of distribution functions S which are flat outside of $\bigcup_{j=1}^m \lfloor p_j, q_j]$. Thus, the search for the MLE of the function S can be restricted to these classes and reduces to maximizing

$$L = \prod_{i=1}^{n} \left(\sum_{j=1}^{m} \alpha_{ij} s_j \right),$$

where

$$\alpha_{ij} = \begin{cases} 1 \text{ if } \lfloor p_j, q_j] \subset \lfloor l_i, u_i] \\ 0 \text{ otherwise.} \end{cases}$$

Therefore, the NPMLE of S can be estimated by constrained maximization of the likelihood L with linear constraints

$$1 - \sum_{j=1}^{m} s_j = 0,$$

$$s_j \geq 0 \quad (j = 1, \ldots, m).$$

This can be accomplished with a variety of algorithms such as the self-consistency algorithm of Turnbull (1976) which can be regarded as an application of

Doubly interval censored data

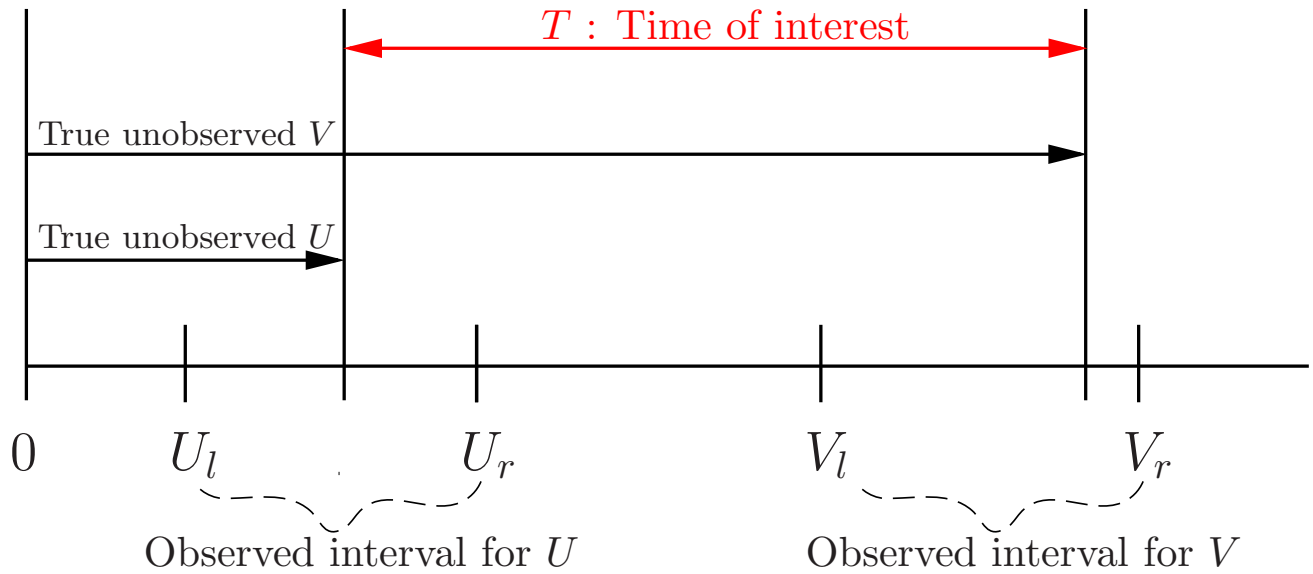
We observe:

- Time to CCYR$= U$: interval censored
- Time to MMR $= V > U$ and interval censored
- Time from CCYR to MMR $= T = V - U$ (Gap time)

Doubly interval censored data

We observe:

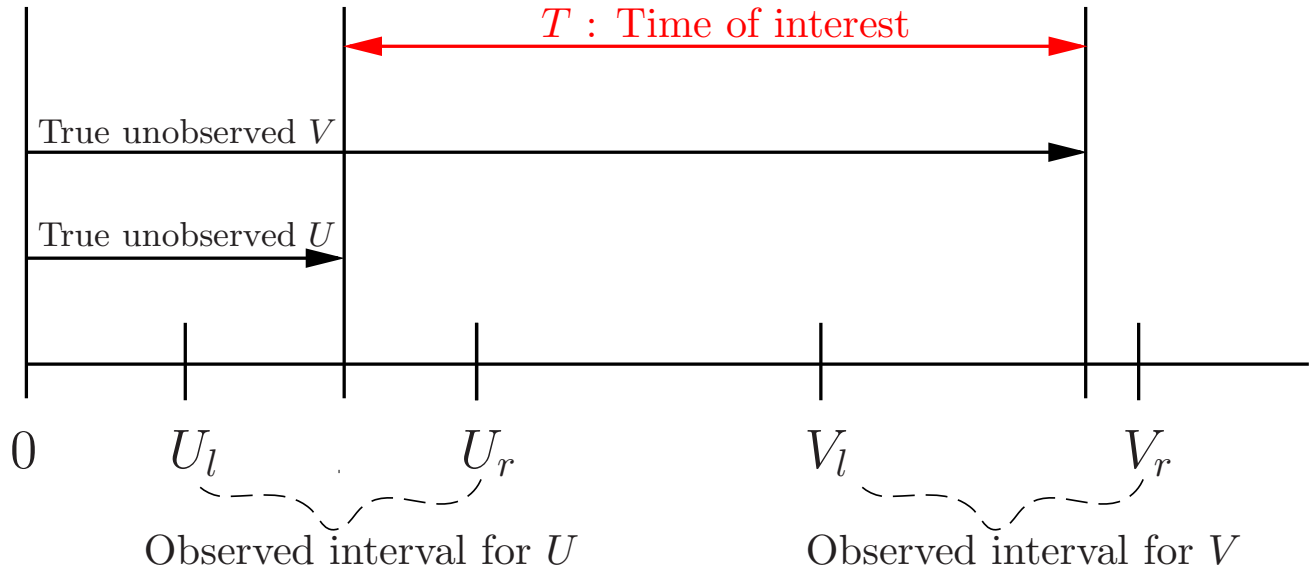
- Time to CCYR= U: interval censored
- Time to MMR = $V > U$ and interval censored
- Time from CCYR to MMR = $T = V - U$ (Gap time)



Doubly interval censored data

We observe:

- Time to CCYR= U: interval censored
- Time to MMR = $V > U$ and interval censored
- Time from CCYR to MMR = $T = V - U$ (Gap time)



- Assume non-informative censoring of U and V
- Double interval censored $\Leftrightarrow$ doubly censored

Doubly interval censored data

Areas of application

- **HIV research**: time between onset of HIV to onset of AIDS
- **Dental research**: time between emergence of tooth to caries
- **Emergency medicine**: time that kidney function deteriorates (below critical level) to time that kidney function recovers
- **Oncology** (example here): but interval censoring must often ignored

Doubly interval censored data

Purpose:

Estimate the distribution (with covariates) of T

under independence of U and T

Current Methods
Reduced likelihood methods

1) Reduced Likelihood methods

- Methods reduce the problem to single/right interval censoring by transforming the data
- Based on single interval censored methods
- Ignoring distribution of U

Reduced likelihood methods

- Maximal interval
  - Compute $\tilde{t}_{li} = v_{li} - u_{ri}$ and $\tilde{t}_{ri} = v_{ri} - u_{li}$
  - Apply interval censored data methods (Turnbull/Pan PH) on $[\tilde{t}_{li}, \tilde{t}_{ri}]$

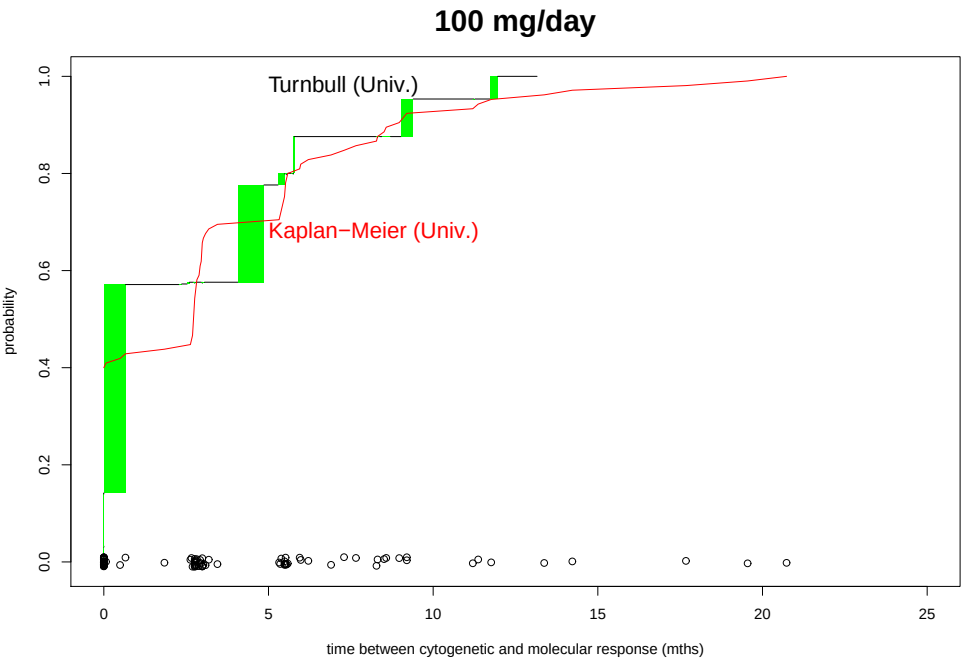
Reduced likelihood methods

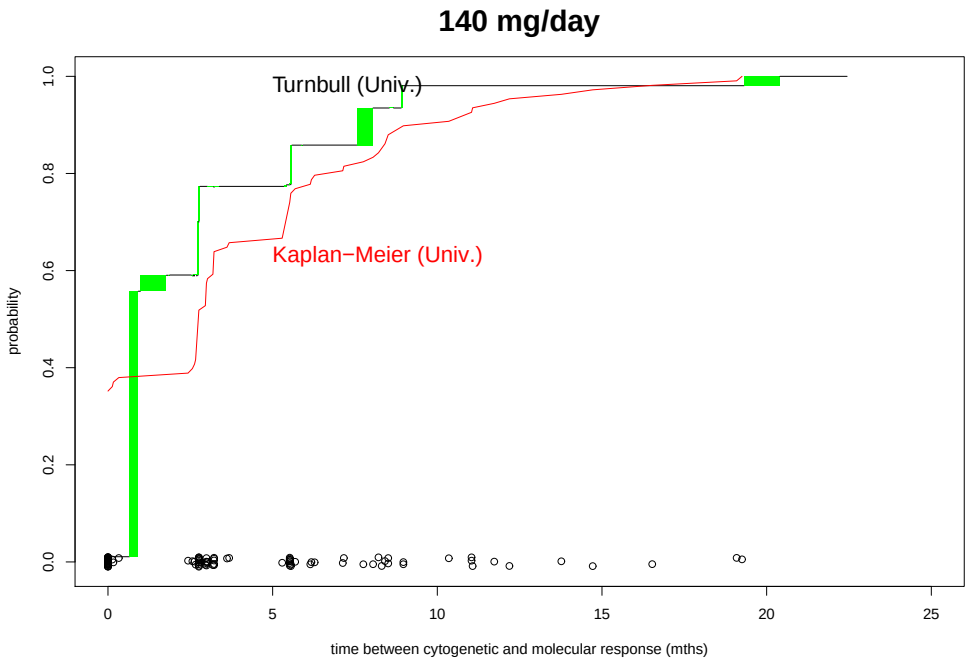
- Maximal interval
  - Compute $\tilde{t}_{li} = v_{li} - u_{ri}$ and $\tilde{t}_{ri} = v_{ri} - u_{li}$
  - Apply interval censored data methods (Turnbull/Pan PH) on $[\tilde{t}_{li}, \tilde{t}_{ri}]$
- Midpoints for U and V
  - Compute $\tilde{u}_i = \frac{u_{li}+u_{ri}}{2}$ and $\tilde{v}_i = \frac{v_{li}+v_{ri}}{2}$
  - Compute $\tilde{t}_i = \tilde{v}_i - \tilde{u}_i$
  - Apply right-censored data methods (KM/Cox PH) on $\tilde{t}_i$

Reduced likelihood methods

- Maximal interval
  - Compute $\tilde{t}_{li} = v_{li} - u_{ri}$ and $\tilde{t}_{ri} = v_{ri} - u_{li}$
  - Apply interval censored data methods (Turnbull/Pan PH) on $[\tilde{t}_{li}, \tilde{t}_{ri}]$
- Midpoints for U and V
  - Compute $\tilde{u}_i = \frac{u_{li}+u_{ri}}{2}$ and $\tilde{v}_i = \frac{v_{li}+v_{ri}}{2}$
  - Compute $\tilde{t}_i = \tilde{v}_i - \tilde{u}_i$
  - Apply right-censored data methods (KM/Cox PH) on $\tilde{t}_i$
- Midpoints for U
  - Compute $\tilde{u}_i = \frac{u_{li}+u_{ri}}{2}$
  - Compute intervals $\tilde{t}_{li} = v_{li} - \tilde{u}_i$ and $\tilde{t}_{ri} = v_{ri} - \tilde{u}_i$
  - Apply interval censored data methods (Turnbull/Pan PH) on $[\tilde{t}_{li}, \tilde{t}_{ti}]$

Example





- KM based on right endpoint of $[v_{ri} - u_{ri}]$
- Turnbull based on $[v_{li} - u_{ri}, v_{ri} - u_{li}]$

Reduced likelihood methods

When are these methods acceptable?

Parameter estimates:

1. When intervals for U and T are small
2. When T stochastically larger than U: $F_U(x) << F_T(x)$
   - True for mid point U and MI methods
   - Reduced likelihood and full likelihood close for $F_U, F_T \sim \exp(\lambda)$
   - Simulations show that it can be generalized to all distributions

Reduced likelihood methods

When are these methods acceptable?

Parameter estimates:

1. When intervals for U and T are small
2. When T stochastically larger than U: $F_U(x) << F_T(x)$
   - True for mid point U and MI methods
   - Reduced likelihood and full likelihood close for $F_U, F_T \sim \exp(\lambda)$
   - Simulations show that it can be generalized to all distributions

Variability of parameter estimate:

- Underestimation of variability of basic data
- Standard errors of parameter estimates underestimated

Reduced likelihood methods

Scenario	Estimator	Mean integrated error^2	Mean integrated bias^2 * 10^4
1	Full likelihood - DeG	0.02	118
T not	Maximal interval	0.08	211
stoch. greater	Mid point for U	0.12	865
	Mid-point for U and V	0.43	4254
2	Full likelihood - DeG	0.01	66
T stoch.	Maximal interval	0.03	61.15
much greater	Mid point for U	0.03	102
	Mid-point for U and V	0.15	1461
3	Full likelihood - DeG	0.24	4582.99
Very small	Maximal interval	0.01	10
intervals	Mid point for U	0.01	2
	Mid-point for U and V	0.01	28

- De Gruttola: reference (Full Likelihood method) - discussed later

Reduced likelihood methods

Scenario	Estimator	Mean integrated error^2	Mean integrated bias^2 * 10^4
1	Full likelihood - DeG	0.02	118
T not	Maximal interval	0.08	211
stoch. greater	Mid point for U	0.12	865
	Mid-point for U and V	0.43	4254
2	Full likelihood - DeG	0.01	66
T stoch.	Maximal interval	0.03	61.15
much greater	Mid point for U	0.03	102
	Mid-point for U and V	0.15	1461
3	Full likelihood - DeG	0.24	4582.99
Very small	Maximal interval	0.01	10
intervals	Mid point for U	0.01	2
	Mid-point for U and V	0.01	28

- De Gruttola: reference (Full Likelihood method) - discussed later
- $F_U << F_T$ implies acceptable estimation by MI and mid point U
  For mid point U and V not much improved

Reduced likelihood methods

Scenario	Estimator	Mean integrated error^2	Mean integrated bias^2 * 10^4
1	Full likelihood - DeG	0.02	118
T not	Maximal interval	0.08	211
stoch. greater	Mid point for U	0.12	865
	Mid-point for U and V	0.43	4254
2	Full likelihood - DeG	0.01	66
T stoch.	Maximal interval	0.03	61.15
much greater	Mid point for U	0.03	102
	Mid-point for U and V	0.15	1461
3	Full likelihood - DeG	0.24	4582.99
Very small	Maximal interval	0.01	10
intervals	Mid point for U	0.01	2
	Mid-point for U and V	0.01	28

- De Gruttola: reference (Full Likelihood method) - discussed later
- $F_U << F_T$ implies acceptable estimation by MI and mid point U
  For mid point U and V not much improved
- Small intervals implies acceptable estimation by reduced likelihood methods

Reduced likelihood methods

Scenario	Estimator	Mean integrated error^2	Mean integrated bias^2 * 10^4
1	Full likelihood - DeG	0.02	118
T not	Maximal interval	0.08	211
stoch. greater	Mid point for U	0.12	865
	Mid-point for U and V	0.43	4254
2	Full likelihood - DeG	0.01	66
T stoch.	Maximal interval	0.03	61.15
much greater	Mid point for U	0.03	102
	Mid-point for U and V	0.15	1461
3	Full likelihood - DeG	0.24	4582.99
Very small	Maximal interval	0.01	10
intervals	Mid point for U	0.01	2
	Mid-point for U and V	0.01	28

- De Gruttola: reference (Full Likelihood method) - discussed later
- $F_U << F_T$ implies acceptable estimation by MI and mid point U
  For mid point U and V not much improved
- Small intervals implies acceptable estimation by reduced likelihood methods
- Identifiability problems with De Gruttola on small intervals

Current Methods
Full likelihood methods

2) Full Likelihood methods

- Methods taking distribution of U into account
- Assume independence between U and T

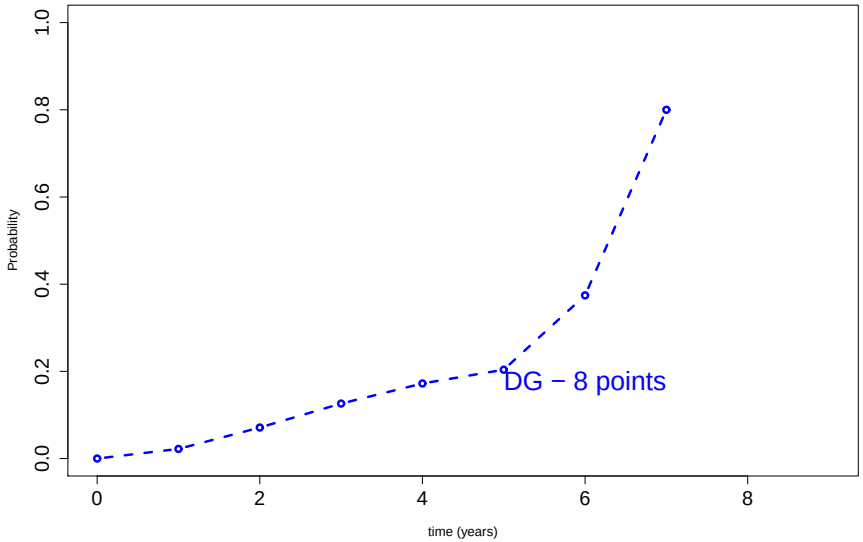
2 types of methods:

1. Based on discretization of the distribution
2. Assuming continuous distributions

Full likelihood methods

a) Methods based on discretization of the distribution

- De Gruttola et al. (1989): **nonparametric approach**
  - No covariates
  - Predefined mass points for U and V
  - Maximizes the likelihood to obtain mass at each (and ONLY) mass points, the rest is **our own imagination**
  - EM (self-consistency) algorithm iterating between estimation of marginal distribution of U and T
  - **Problems**: convergence + identifiability



Full likelihood methods

a) Methods based on discretization of the distribution

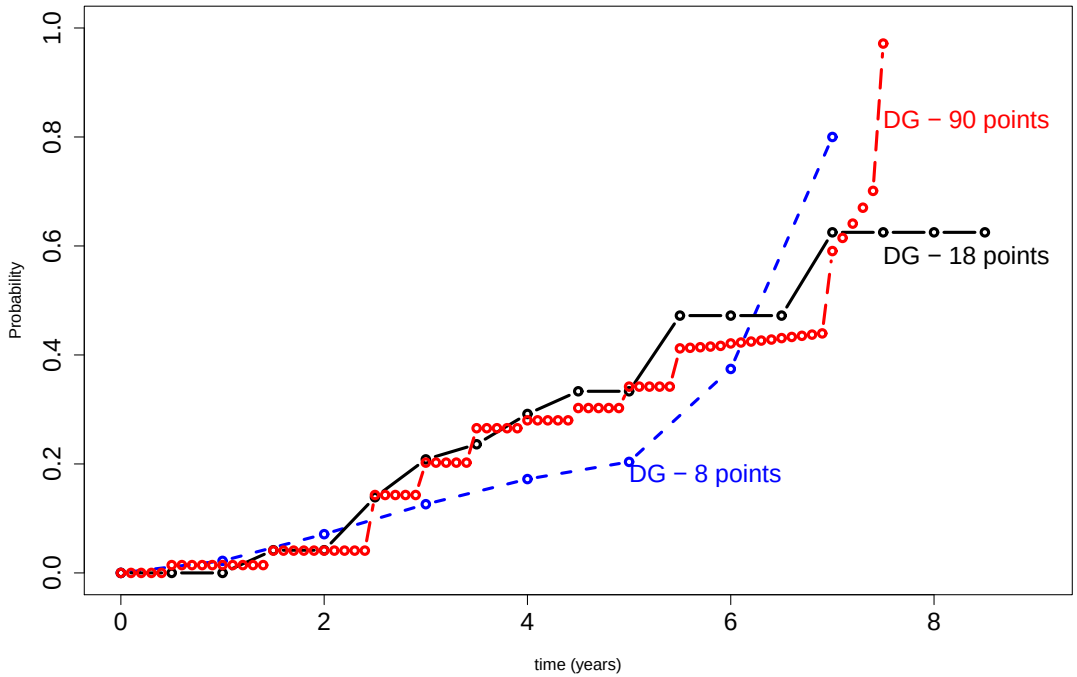
- Kim et al. (1993)
  - Extension of DeG to covariates
  - Assuming Cox PH model
  - Estimation by self consistency algorithm and direct maximization of likelihood

Both approaches (DeG + Kim)

Influenced by chosen/prespecified locations of the mass points

Full likelihood methods
Issues with methods based on discretization

Not enough or too many points of mass $\rightarrow$ bias



Example from De Gruttola et al. (1986) - 18 mass points

Full likelihood methods

b) Methods for continuous distributions

- Gomez et al. (1999): conditional ML approach without covariates (not fully efficient)
- Sun et al (1999): profile likelihood approach with covariates

Issues:

1. Based on right censored V
2. Not suitable for overlapping intervals

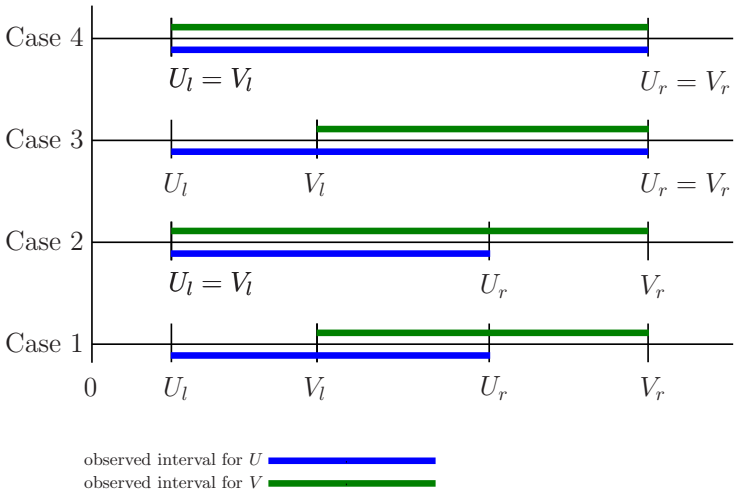
Full likelihood methods

b) Methods for continuous distributions

- Gomez et al. (1999): conditional ML approach without covariates (not fully efficient)
- Sun et al (1999): profile likelihood approach with covariates

Issues:

1. Based on right censored V
2. Not suitable for overlapping intervals



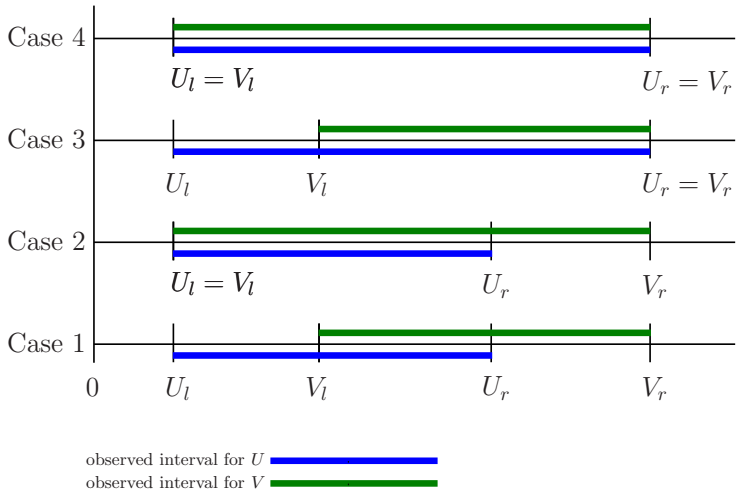
Full likelihood methods

b) Methods for continuous distributions

- Gomez et al. (1999): conditional ML approach without covariates (not fully efficient)
- Sun et al (1999): profile likelihood approach with covariates

Issues:

1. Based on right censored V
2. Not suitable for overlapping intervals



- Here 50% overlapping
- Unidentifiability when too often

Current Methods

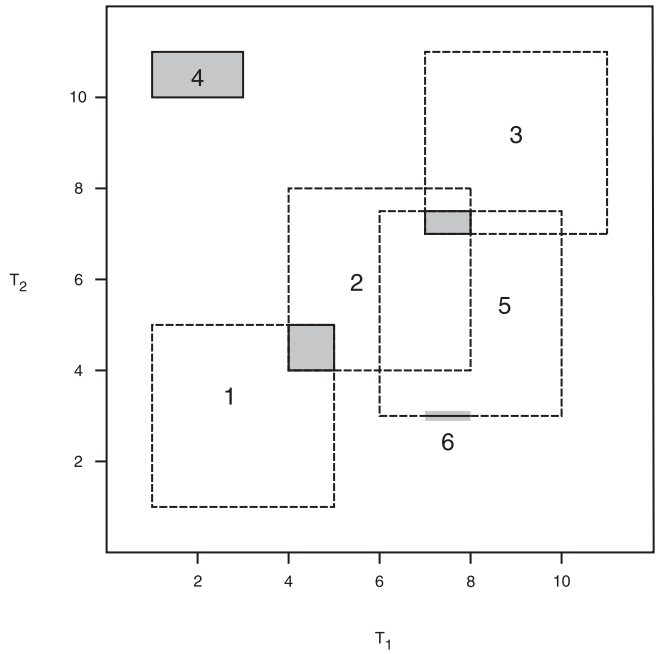
3) Other approaches

- Parametric methods
- Bayesian semi-parametric methods
- Extension to bivariate NPMLE

- **Parametric**: mispecification issues
- **Bayesian semi-parametric methods**:
  - Fitting AFT by a flexible mixture of normals using Bayesian approach (Komárek & Lesaffre, 2006,2007,2008)
  - Poisson Dirichlet process (Jara et al, 2010) (Bayesian approach)

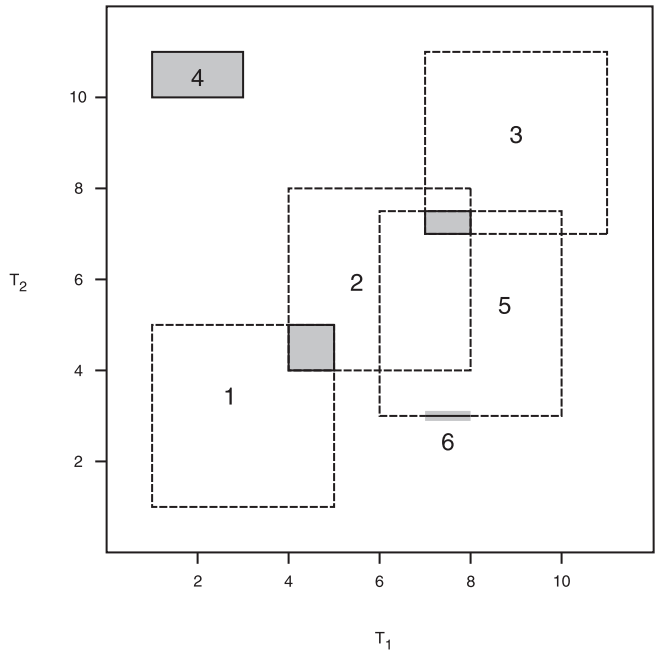
Current Methods

- **Extension to bivariate NPMLE?**



Current Methods

- **Extension to bivariate NPMLE?**



- **Problems**
  - Not clear how to deal with $U \leq V$
  - Analysis should be done in U and T
  - How to include covariates?

Conclusions so far

- Difficult to tackle the problem of DI data in a bivariate manner
- All approaches work in 2 steps
- For DI data all approaches assume pre-chosen mass points

StEM

Motivation

In a clinical trial context:

- Need for a more formal approach
- Independent of prespecified mass points to avoid subjectivity in estimate
- Allowing for overlap of intervals

StEM

Outline:

- Introduction the concept - Based on EM
- Justification of the need for iterative algorithm
- Description of algorithm

Concept

We have:

- U (time to CCYR) is interval censored
- V (time to MMR) is interval censored

Exact event times of U and V unknown (missing)

- EM algorithm provides MLE in presence of missing data
- Estimating the distribution of right censored data: KM estimator
- KM is maximum likelihood

- Assuming only U unknown + interval censoring is **not better**

Concept

- Observed data $= [u_{li}, u_{ri}], [v_{li}, v_{ri}]$ $(i = 1, \ldots, n)$
- $V = U + T$
- $F_T(t)$ given by KM likelihoods based on data t_i: $L(p_i|t_i)$ $(i = 1, \ldots, n)$
- Parameters $p_i =$ KM mass at death time t_i to be estimated

Concept

- Observed data $= [u_{li}, u_{ri}], [v_{li}, v_{ri}]$ $(i = 1, \ldots, n)$
- $V = U + T$
- $F_T(t)$ given by KM likelihoods based on data t_i: $L(p_i|t_i)$ $(i = 1, \ldots, n)$
- Parameters $p_i =$ KM mass at death time t_i to be estimated

EM algorithm (subindex i runs over all observations)

Concept

- Observed data $= [u_{li}, u_{ri}], [v_{li}, v_{ri}]$ $(i = 1, \ldots, n)$
- $V = U + T$
- $F_T(t)$ given by KM likelihoods based on data t_i: $L(p_i|t_i)$ $(i = 1, \ldots, n)$
- Parameters $p_i =$ KM mass at death time t_i to be estimated

EM algorithm (subindex i runs over all observations)

E-step: compute the expected likelihood over missing data (U, T), given the observed data, given the estimate at previous iteration
At iteration k

$$Q^k(p_i, p_i^{k-1}) = \mathrm{E}_{u_i,t_i}[\log L(p_i|t_i)|\text{observed data}, p_i^{k-1}]$$
$$= \int_{u_i,t_i} \log L(p_i|t_i) dF(u_i, t_i|[u_{li}, u_{ri}][v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

Concept

- Observed data $= [u_{li}, u_{ri}], [v_{li}, v_{ri}]$ $(i = 1, \ldots, n)$
- $V = U + T$
- $F_T(t)$ given by KM likelihoods based on data t_i: $L(p_i|t_i)$ $(i = 1, \ldots, n)$
- Parameters $p_i =$ KM mass at death time t_i to be estimated

EM algorithm (subindex i runs over all observations)

E-step: compute the expected likelihood over missing data (U, T), given the observed data, given the estimate at previous iteration
At iteration k

$$Q^k(p_i, p_i^{k-1}) = \mathrm{E}_{u_i,t_i}[\log L(p_i|t_i)|\text{observed data}, p_i^{k-1}]$$
$$= \int_{u_i,t_i} \log L(p_i|t_i) dF(u_i, t_i|[u_{li}, u_{ri}][v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

M-step: maximize Q_k to obtain $\hat{F}_T^k(.)$

Concept

Conditional distribution of missing data given observed data:

$$F(u, t \,|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$
$$= F_T(t \mid u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) \; F_U(u \,|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

- By independence of U and T
- Note: $F_T(t \mid u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \hat{F}_T^{k-1}(.)$
- $\hat{F}_T$ and $\hat{F}_U$ are updated in EM iterations
- Marginal F_U estimated ONCE on $[u_{li}, u_{ri}]$ (Turnbull)

Concept

Conditional distribution of missing data given observed data:

$$F(u, t \,|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$
$$= F_T(t \mid u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) \; F_U(u \,|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

- By independence of U and T
- Note: $F_T(t \mid u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \hat{F}_T^{k-1}(.)$
- $\hat{F}_T$ and $\hat{F}_U$ are updated in EM iterations
- Marginal F_U estimated ONCE on $[u_{li}, u_{ri}]$ (Turnbull)

$\Rightarrow$ Conditional density of U given data depends on V and $\hat{F}_T^{k-1}(.)$

$\Rightarrow$ No closed expression for $Q^k(p_i, p_i^{k-1})$

Concept

Conditional distribution of missing data given observed data:

$$F(u, t \,|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$
$$= F_T(t \mid u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) \; F_U(u \,|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

- By independence of U and T
- Note: $F_T(t \mid u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \hat{F}_T^{k-1}(.)$
- $\hat{F}_T$ and $\hat{F}_U$ are updated in EM iterations
- Marginal F_U estimated ONCE on $[u_{li}, u_{ri}]$ (Turnbull)

$\Rightarrow$ Conditional density of U given data depends on V and $\hat{F}_T^{k-1}(.)$

$\Rightarrow$ No closed expression for $Q^k(p_i, p_i^{k-1})$

$\Rightarrow$ **Iterative algorithm** needed

Concept

Q^k has no closed form $\Rightarrow$ use Stochastic EM algorithm

- **Replace integration** on U and T by generating
  - $(\bar{u}_1, \ldots, \bar{u}_n)_q$ out of

    $$\hat{F}_U(u|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

  - and $(\bar{t}_1, \ldots, \bar{t}_n)_q$ from

    $$\hat{F}_T(t|u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.))$$

  for $q = 1, \ldots, m$ (m generated datasets)

- **Maximize** m likelihoods and average over m estimates to obtain updated $\hat{F}_T^k(.)$

StEM algorithm

Initialization Estimate $F_T^0(t)$ using mid-point of intervals

StE-step k (1) Generate $(\bar{u}_1, \ldots, \bar{u}_n)_q$ from

$$\hat{F}_U(\ u\ |[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \frac{1}{cst} \int_{u_{li}}^{u} \int_{v_{li}-u}^{v_{ri}-u} d\hat{F}_T^{k-1}(t) d\hat{F}_U(u)$$

where

StEM algorithm

Initialization Estimate $F_T^0(t)$ using mid-point of intervals

StE-step k (1) Generate $(\bar{u}_1, \ldots, \bar{u}_n)_q$ from

$$\hat{F}_U(\ u\ |[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \frac{1}{cst} \int_{u_{li}}^{u} \int_{v_{li}-u}^{v_{ri}-u} d\hat{F}_T^{k-1}(t) d\hat{F}_U(u)$$

where

- $cst = \int_{u_{li}}^{u_{ri}} \int_{v_{li}-u}^{v_{ri}-u} d\hat{F}_T^{k-1}(t) d\hat{F}_U(u)$
- $F_U(u)$ estimated by Turnbull

StEM algorithm

StE-step k (2) Generate $(\bar{t}_1, \ldots, \bar{t}_n)_q$ from

$$\hat{F}_T(\ t\ |u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \int_{v_{li}-u}^{t} d\hat{F}_T^{k-1}(t) / \int_{v_{li}-u}^{v_{ri}-u} d\hat{F}_T^{k-1}(t)$$

StEM algorithm

StE-step k (2) Generate $(\bar{t}_1, \ldots, \bar{t}_n)_q$ from

$$\hat{F}_T(\ t\ |u, [v_{li}, v_{ri}], \hat{F}_T^{k-1}(.)) = \int_{v_{li}-u}^{t} d\hat{F}_T^{k-1}(t) / \int_{v_{li}-u}^{v_{ri}-u} d\hat{F}_T^{k-1}(t)$$

M-step

- Apply KM estimator based on $(\bar{t}_1, \ldots, \bar{t}_n)_q$
  Maximize m likelihoods to obtain m estimates of F_T
- Average m estimates to obtain $\hat{F}_T^k(.)$

Repeat StE-step and M-step until convergence.

StEM algorithm: Variance calculation

StEM Variance estimate has 2 components:

- Average of m within-variances
- Between-iteration variance

Formally, at a fixed time t, at final iteration k:

$$\hat{\sigma}^2(t) = \frac{1}{m}\sum_{q=1}^{m}\hat{\sigma}_q^2(t) + (1+\frac{1}{m})\sum_{q=1}^{m}(\hat{F}_T^k(t)_q - \hat{F}_T^k(t))^2$$

where

- $\hat{F}_T^k(t)_q$ is the KM estimate on $(\bar{t}_1, \ldots, \bar{t}_n)_q$ at iteration k
- $\hat{\sigma}_q^2(t)$ is the KM estimate of the variance at time t of $\hat{F}_T^k(t)_q$

StEM algorithm: Cox PH model

Model

$$\begin{aligned} U &\sim F_U(u) \\ T &\sim S_T(t|X) = S_T(t|X=0)^{e^{\beta X}} \end{aligned}$$

- Covariate acting on F_T only
- Notation: "baseline distribution" $F_{T0}(t) = 1 - S_T(t|X=0)(t)$

StEM algorithm: Cox PH model

StE-step k (1) Generate $(\bar{u}_1, \ldots, \bar{u}_n)_q$ from

$$\hat{F}_U(u|[u_{li}, u_{ri}], [v_{li}, v_{ri}], \hat{F}_{T0}^{k-1}(.|X), \beta^k)$$

StE-step k (2) Generate $(\bar{t}_1, \ldots, \bar{t}_n)_q$ from

$$\hat{F}_T(t|X, u, [v_{li}, v_{ri}], \hat{F}_{T0}^{k-1}(.), \beta^k)$$

M-step

- Apply Cox PH estimator based on $(\bar{t}_1, \ldots, \bar{t}_n)_q$
  Maximize m likelihoods to obtain m estimates of F_{T0} and β
- Average m estimates to obtain $\hat{F}_{T0}^k(.)$ and $\hat{\beta}^k$

Repeat StE-step and M-step until convergence.

Simulations -Estimation of Distribution

Scen-ario	Distributions	Estimator	$\hat{F}(p_{50})$	(std)	(ste)	Mean integrated bias^2	MISE
1	$U \approx \exp(\lambda = 10)$	Max int Turnbull	0.54	0.08	0.09	0.0044	0.0275
	$T \approx \exp(\lambda = 0.5)$	De Gruttola (Perc)	0.52	0.06	0.07	0.0010	0.0136
		De Gruttola (Sun)	0.51	0.06	0.07	0.0008	0.0129
		StEM	0.50	0.08	0.08	0.0005	0.0161
2	$U \approx$ Weibull(shape=2,scale=5)	Max int Turnbull	0.42	0.18	0.24	0.0167	0.0551
	$T \approx$ log normal(mean=2,sd=0.1)	De Gruttola (Perc)	0.54	0.08	0.08	0.0190	0.032
		De Gruttola (Sun)	0.55	0.11	0.11	0.5093	0.5568
		StEM	0.47	0.08	0.11	0.0085	0.0317
3	$U \approx$ Weibull(shape=1.7,scale=.83)	Max int Turnbull	0.42	0.16	0.23	0.0114	0.0415
	$T \approx$ log normal(mean=2.5,sd=0.05)	De Gruttola (Perc)	0.56	0.08	0.08	0.0695	0.089
		De Gruttola (Sun)	0.56	0.11	0.11	1.2173	1.2860
		StEM	0.50	0.08	0.11	0.0375	0.0687
4	$U \approx$ Weibull(shape=1.7,scale=.83)	Max int Turnbull	0.54	0.11	0.11	0.0459	0.1134
	$T \approx$ Weibull(shape=1.7,scale=.83)	De Gruttola (Perc)	0.53	0.06	0.06	0.0058	0.0342
		De Gruttola (Sun)	0.54	0.07	0.07	0.0240	0.0555
		StEM	0.50	0.08	0.07	0.0008	0.038

- Bias better, MISE better / similar to De Gruttola estimator

Simulations -Estimation of Distribution

Scen-ario	Distributions	Estimator	$\hat{F}(p_{50})$	(std)	(ste)	Mean integrated bias^2	MISE
1	$U \approx \exp(\lambda = 10)$	Max int Turnbull	0.54	0.08	0.09	0.0044	0.0275
	$T \approx \exp(\lambda = 0.5)$	De Gruttola (Perc)	0.52	0.06	0.07	0.0010	0.0136
		De Gruttola (Sun)	0.51	0.06	0.07	0.0008	0.0129
		StEM	0.50	0.08	0.08	0.0005	0.0161
2	$U \approx$ Weibull(shape=2,scale=5)	Max int Turnbull	0.42	0.18	0.24	0.0167	0.0551
	$T \approx$ log normal(mean=2,sd=0.1)	De Gruttola (Perc)	0.54	0.08	0.08	0.0190	0.0329
		De Gruttola (Sun)	0.55	0.11	0.11	0.5093	0.5568
		StEM	0.47	0.08	0.11	0.0085	0.0317
3	$U \approx$ Weibull(shape=1.7,scale=.83)	Max int Turnbull	0.42	0.16	0.23	0.0114	0.0415
	$T \approx$ log normal(mean=2.5,sd=0.05)	De Gruttola (Perc)	0.56	0.08	0.08	0.0695	0.0899
		De Gruttola (Sun)	0.56	0.11	0.11	1.2173	1.2860
		StEM	0.50	0.08	0.11	0.0375	0.0687
4	$U \approx$ Weibull(shape=1.7,scale=.83)	Max int Turnbull	0.54	0.11	0.11	0.0459	0.1134
	$T \approx$ Weibull(shape=1.7,scale=.83)	De Gruttola (Perc)	0.53	0.06	0.06	0.0058	0.0342
		De Gruttola (Sun)	0.54	0.07	0.07	0.0240	0.0555
		StEM	0.50	0.08	0.07	0.0008	0.0384

- Bias better, MISE better / similar to De Gruttola estimator
- Better than Turnbull based on $[v_{ri} - u_{li}, v_{li} - u_{ri}]$

Simulations -Estimation of Distribution

Scen-ario	Distributions	Estimator	$\hat{F}(p_{50})$	(std)	(ste)	Mean integrated bias^2	MISE
1	$U \approx \exp(\lambda = 10)$	Max int Turnbull	0.54	0.08	0.09	0.0044	0.0275
	$T \approx \exp(\lambda = 0.5)$	De Gruttola (Perc)	0.52	0.06	0.07	0.0010	0.0136
		De Gruttola (Sun)	0.51	0.06	0.07	0.0008	0.0129
		StEM	0.50	0.08	0.08	0.0005	0.0161
2	$U \approx$ Weibull(shape=2,scale=5)	Max int Turnbull	0.42	0.18	0.24	0.0167	0.0551
	$T \approx$ log normal(mean=2,sd=0.1)	De Gruttola (Perc)	0.54	0.08	0.08	0.0190	0.032
		De Gruttola (Sun)	0.55	0.11	0.11	0.5093	0.5568
		StEM	0.47	0.08	0.11	0.0085	0.0317
3	$U \approx$ Weibull(shape=1.7,scale=.83)	Max int Turnbull	0.42	0.16	0.23	0.0114	0.0415
	$T \approx$ log normal(mean=2.5,sd=0.05)	De Gruttola (Perc)	0.56	0.08	0.08	0.0695	0.089
		De Gruttola (Sun)	0.56	0.11	0.11	1.2173	1.2860
		StEM	0.50	0.08	0.11	0.0375	0.0687
4	$U \approx$ Weibull(shape=1.7,scale=.83)	Max int Turnbull	0.54	0.11	0.11	0.0459	0.1134
	$T \approx$ Weibull(shape=1.7,scale=.83)	De Gruttola (Perc)	0.53	0.06	0.06	0.0058	0.0342
		De Gruttola (Sun)	0.54	0.07	0.07	0.0240	0.0555
		StEM	0.50	0.08	0.07	0.0008	0.038

- Bias better, MISE better / similar to De Gruttola estimator
- Better than Turnbull based on $[v_{ri} - u_{li}, v_{li} - u_{ri}]$
- Large influence of prespecified mass points on De Gruttola estimator

Simulations - Cox PH model

- $U \sim \exp(1)$ and $T \sim$ Weibull(1.7, 5.83)
- Number of generated values: 20 for first 50 iterations, 100 after

Scenario	β	Estimator	$\hat{\beta}$	95% coverage Probability	Power
1	0.5	Mid point	0.07	0.16	0.08
		Pan (Univ.)	0.58	0.81	0.79
		StEM	0.52	0.91	0.80
2	-0.5	Mid point	-0.34	0.84	0.72
		Pan (Univ.)	-0.58	0.9	0.82
		StEM	-0.53	0.9	0.88
4	-0.2	Mid point	-0.13	0.95	0.13
		Pan (Univ.)	-0.22	0.94	0.23
		StEM	-0.22	0.97	0.23
5	0	Mid point	0.02	0.88	0.12
		Pan (Univ.)	0.04	0.92	0.08
		StEM	0.01	0.95	0.05
6(X cont.)	-0.5	StEM	-0.51	0.94	0.62

- Univariate estimator not good

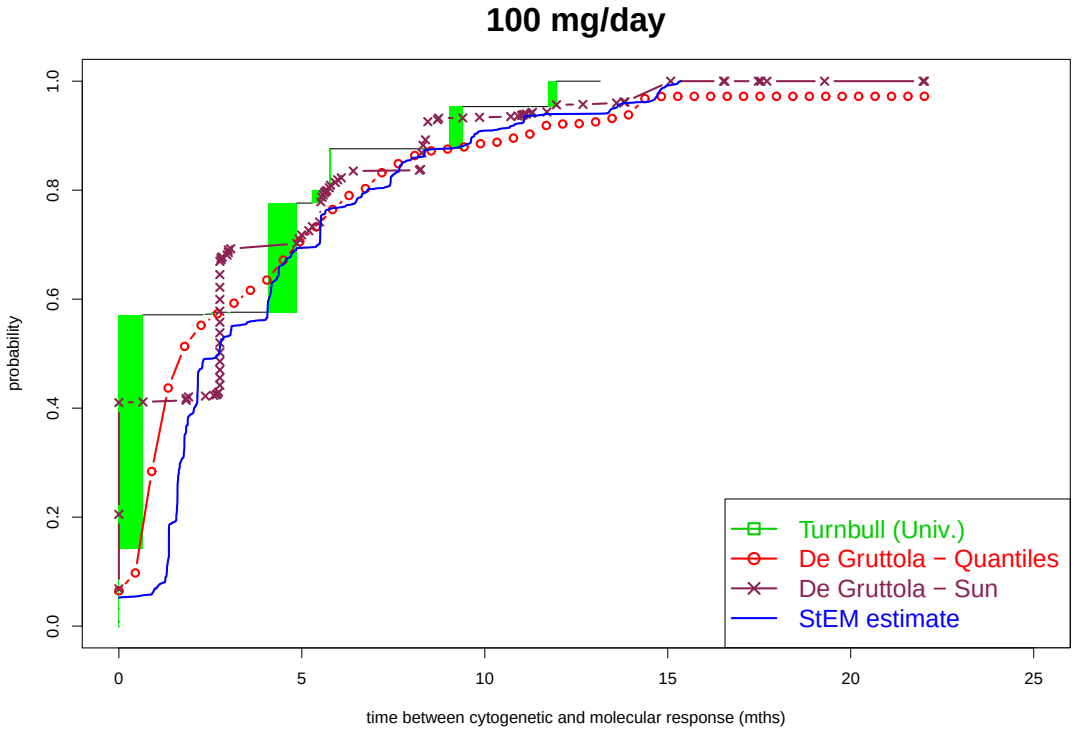
Simulations - Cox PH model

- $U \sim \exp(1)$ and $T \sim$ Weibull(1.7, 5.83)
- Number of generated values: 20 for first 50 iterations, 100 after

Scenario	β	Estimator	$\hat{\beta}$	95% coverage Probability	Power
1	0.5	Mid point	0.07	0.16	0.08
		Pan (Univ.)	0.58	0.81	0.79
		StEM	0.52	0.91	0.80
2	-0.5	Mid point	-0.34	0.84	0.72
		Pan (Univ.)	-0.58	0.9	0.82
		StEM	-0.53	0.9	0.88
4	-0.2	Mid point	-0.13	0.95	0.13
		Pan (Univ.)	-0.22	0.94	0.23
		StEM	-0.22	0.97	0.23
5	0	Mid point	0.02	0.88	0.12
		Pan (Univ.)	0.04	0.92	0.08
		StEM	0.01	0.95	0.05
6(X cont.)	-0.5	StEM	-0.51	0.94	0.62

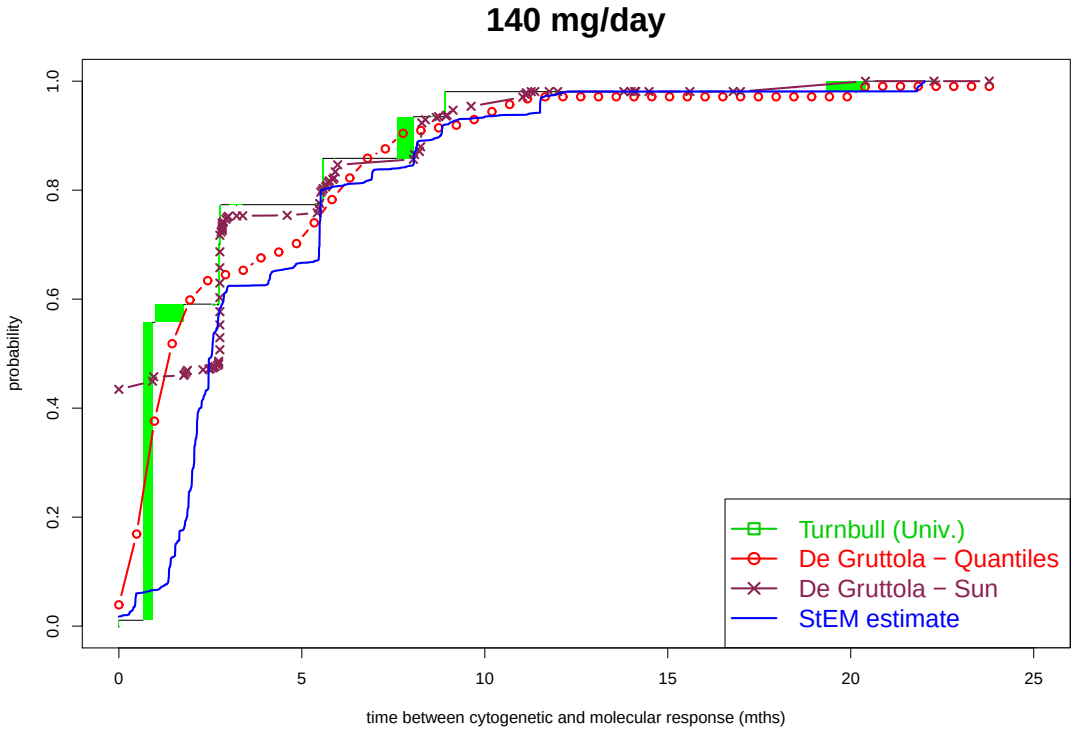
- Univariate estimator not good
- Shows acceptable bias and coverage probability

Motivating example: time between CCyR and MMR



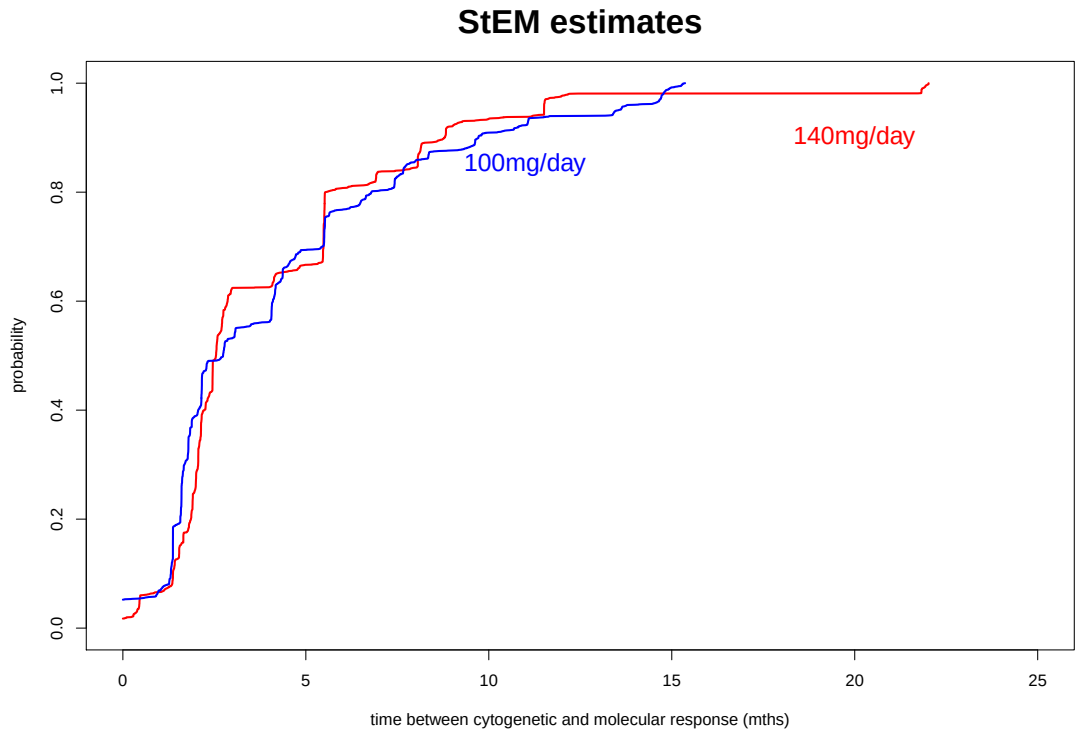
- Estimated distribution by StEM different from De Gruttola and Turnbull (based only on T)

Motivating example: time between CCyR and MMR



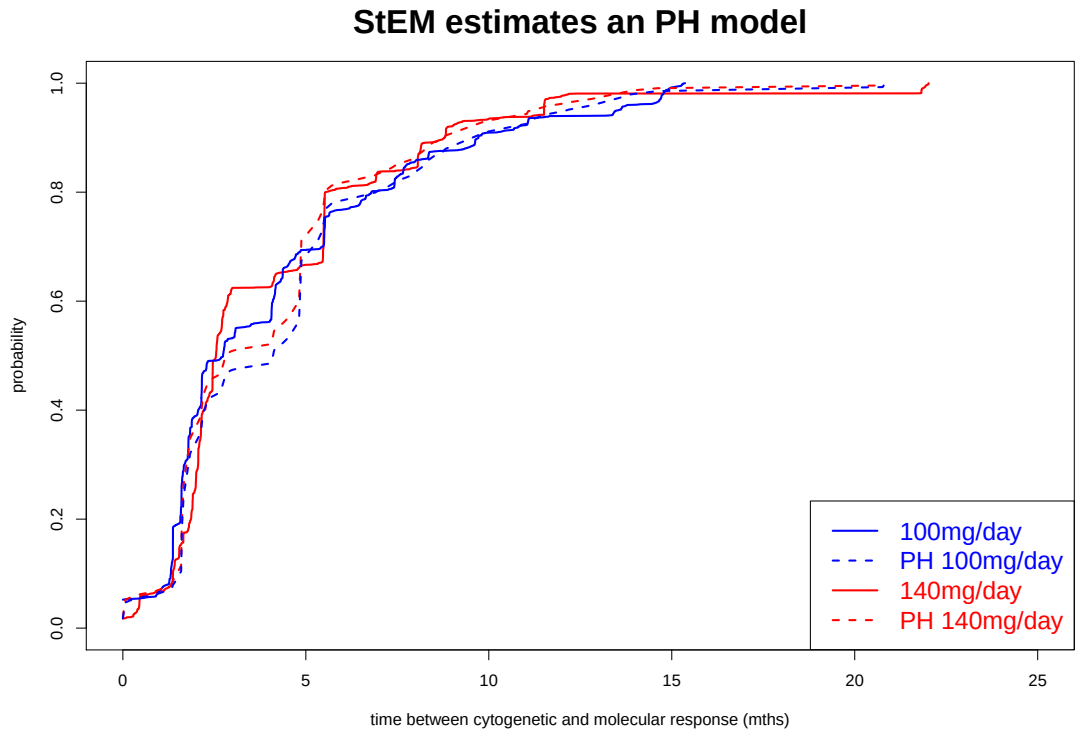
- Estimated distribution by StEM different from De Gruttola and Turnbull (based only on T)

Motivating example: time between CCyR and MMR



- Estimated distribution by StEM different from De Gruttola and Turnbull (based only on T)
- No difference between doses from separate estimation

Motivating example: time between CCyR and MMR



- Estimated distribution by StEM different from De Gruttola and Turnbull (based only on T)
- No difference between doses from separate estimation
- No difference between doses from Cox PH model: $\beta = -0.10(0.14)$

Discussion

- **Reduced Likelihood methods**
  - Simple to implement (existing software available)
  - Work well under some conditions (small intervals/ T stochastically larger U)

Discussion

- **Reduced Likelihood methods**
  - Simple to implement (existing software available)
  - Work well under some conditions (small intervals/ T stochastically larger U)
- **Full Likelihood Methods**
  - Based on discretizing the distribution
    Issues with choice of location and number of mass points
  - Continuous methods
    Restricted to right censored V
    Do not allow overlapping

Discussion

- **Reduced Likelihood methods**
  - Simple to implement (existing software available)
  - Work well under some conditions (small intervals/ T stochastically larger U)
- **Full Likelihood Methods**
  - Based on discretizing the distribution
    Issues with choice of location and number of mass points
  - Continuous methods
    Restricted to right censored V
    Do not allow overlapping
- **StEM**
  - Not impacted by prespecified mass points
  - Allows overlapping / interval censored V
  - Similar/better performance compared to full likelihood methods

The method is implemented in **R** by second author

Thank you for your attention!

Publicity


5th International Meeting on
Methodological Issues In Oral Health Research
ÖGZMK
STEIERMARK
10th – 14th of April 2012, Graz
Welcome | General Information | Scientific Program | Registration | Venue | Important Dates
Welcome to the
5th International Meeting:
Methodological Issues in Oral Health Research
April 10-14, 2012 Graz, Austria
www.dentalbiostat2012.org