

Rank Estimation of Treatment Differences Based on Repeated Measurements Subject to Dependent Censoring

David V. Glidden, UCSF

Feb 21, 2002
Genentech

<http://www.epibiostat.ucsf.edu/dave/>
J. Amer Statist Assoc **94: 888-895.**

Slide 0

Repeated Measurements Data:

Often we want to make inference about the effect of a treatment on repeated measurements data:

- the effect of an intervention on the number of HIV-RNA copies and/or CD4+ cells per mm³, or
- an interventions effect on serum cholesterol, or
- an interventions effect on body weight

Often, we want to compare this to another drug at a series of times after a study begins.

Slide 1

Example: ACTG 019

Three-arm study of zidovudine (ZDV) in HIV infected persons with CD4+ cells mm³ < 500

$$Z_i = \begin{cases} 1: & \mathbf{500mg\ ZDV/day} \\ 0: & \mathbf{Placebo} \end{cases}$$

Primary Endpoint: Progression of clinical disease
Secondary Analysis: Changes in CD4 counts during the study
869 subjects enrolled in the study (421: PLC, 448: ZDV).
The analysis of CD4 data is complicated by the frequent missing observations.

Slide 2

Three Kinds of Missing Data

Study Opens Study Analyzed

Intermittent _____ _____
 t_1 t_3

Late Entry _____
 t_1 t_2

Withdrawal _____
 t_1 t_2 t_3

Slide 3

Notation

Let $X_i = (X_{i1}, \dots, X_{iK})$ denote the CD4 counts;
 $i = 1, \dots, n$

Measured at times (t_1, \dots, t_K)

C_i : time to study termination

D_i : time to dropout

$T_i = \min(C_i, D_i)$: follow-up time

$$\delta_i = \begin{cases} 0 & \text{if } C_i < D_i \\ 1 & \text{if } C_i \geq D_i \end{cases}$$

C_i and D_i compete to censor X_i

The indicator vector $E_i = (\epsilon_{i1}, \dots, \epsilon_{iK})$ indicates whether a value is measured.

Slide 4

A Model for X

Marginally,

$$(X_{ij} - \theta_j | Z_i = 1) \stackrel{D}{=} (X_{ij} | Z_i = 0)$$

θ_j is the treatment effect at time t_j .

Θ can be estimated by solving:

$$\sum_{i:Z_i=0} \sum_{j:Z_j=1} \epsilon_{ik} \epsilon_{jk} \{I(X_{jk} - \theta_k \leq X_{ik}) - 1/2\}$$

$k = 1, \dots, K$.

Slide 5

'Dependent' Censoring

- Assume C_i is independent of D_i and X_i
- If X_i and D_i are correlated, we say that censoring is **dependent**.

Observed Data: $(X_{ij} | D_i > t_j, Z_i)$

Complete Data: $(X_{ij} | Z_i)$

- The observed data is a sample from a conditional distribution which does not always resemble the complete data.

Slide 6

A Joint Model

A **terse** statement of the model:

$$(X_{i1} - \theta_1, \dots, X_{iK} - \theta_K, g(D_i) | Z_i = 1) \stackrel{D}{=} (X_{i1}, \dots, X_{iK}, D_i | Z_i = 0)$$

where $g(\cdot)$ is an unknown, strictly increasing function.

$(X, D | Z = 0)$ has an unspecified distribution,

$\Theta = (\theta_1, \dots, \theta_K)$ indexes the effect of Rx of X ,

$g(\cdot)$ indexes the effect of Rx on D .

Note:

$$g(\cdot) = \Lambda_0^{-1} \Lambda_1(\cdot)$$

where $D | Z = 0 \sim \Lambda_0(\cdot)$ and $D | Z = 1 \sim \Lambda_1(\cdot)$

Slide 7

Conditional Form

Marginally,

$$(X_{ij} - \theta_j | Z_i = 1) \stackrel{D}{=} (X_{ij} | Z_i = 0)$$

Conditionally,

$$(X_{ij} - \theta_j | D_i > g^{-1}(t_j), Z_i = 1) \stackrel{D}{=} (X_{ij} | D_i > t_j, Z_i = 0)$$

Note: The conditional form is a complicated beast which involves both Θ and $g(\cdot)$

$X_i | D_i > t$ is a biased sample, trajectory depends on the joint distribution with magnitude determined by % censored (ie, with $D_i \leq t$)

Slide 8

THE BIG IDEA

Under this model:

$$(X_{ij} - \theta_j | D_i > t, Z_i = 1) \stackrel{D}{=} (X_{ij} | D_i > t^*, Z_i = 0)$$

if

$$P(D_i > t | Z_i = 1) = P(D_i > t^* | Z_i = 0)$$

ie, $\Lambda_1(t) = \Lambda_0(t^*)$

We can find two biased samples that obey the same Θ treatment difference.

Note: $t^* = g(t)$ and $t = g^{-1}(t^*)$

Lets put this property to work!

Slide 9

Estimating Θ with Dependent Censoring

If $g(\cdot)$ is known, in the presence of dependent censoring, we can estimate Θ by solving $S_{1k}(\theta_k, g_k, h_k) = 0$

$$\sum_{i:Z_i=0} \sum_{j:Z_j=1} \epsilon_{ik} \epsilon_{jk} I(D_i > h_k) I(D_j > g_k) \{I(X_{jk} - \theta_k \leq X_{ik}) - 1/2\}$$

where $g_k = g(t_k)$, $h_k = g^{-1}(t_k)$ and $k = 1, \dots, K$.

Unfortunately, $g(\cdot)$ is unknown...

Slide 10

Estimating (g_k, h_k)

The previous estimating equations only depend on $g(\cdot)$ through g_k and h_k , $k = 1, \dots, K$. They can be estimated by solving:

$$S_{2k}(g_k) = \hat{\Lambda}_0(g_k) - \hat{\Lambda}_1(t_k)$$

and

$$S_{3k}(h_k) = \hat{\Lambda}_1(h_k) - \hat{\Lambda}_0(t_k)$$

$k = 1, \dots, K$

$\hat{\Lambda}$ denotes the Nelson-Aalen estimator.

In some ugly situations (g_k, h_k) can't be estimated!

Slide 11

Three Steps for Estimation

1. Estimate \hat{g}_k and \hat{h}_k , $k = 1, \dots, K$
2. Use the above to 'recensor' the data, and
3. Calculate the parameters from the reduced data

Asymptotic theory is largely straightforward.

Standard errors and confidence intervals are based on a resampling method due to Parzen and Wei (1994).

Slide 12

Example: ACTG 019

Three-arm study of zidovudine (ZDV) in HIV infected persons with CD4+ cells $\text{mm}^3 < 500$

$$Z_i = \begin{cases} 1: & \mathbf{500mg\ ZDV/day} \\ 0: & \mathbf{Placebo} \end{cases}$$

Primary Endpoint: Progression of clinical disease

Secondary Analysis: Changes in CD4 counts during the study

869 subjects enrolled in the study (421: PLC, 448: ZDV).

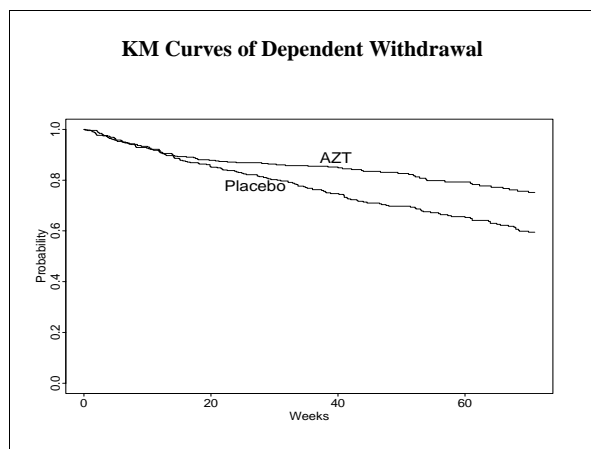
The analysis of CD4 data is complicated by the frequency of withdrawals.

Slide 13

Experience of Volunteers in 019

	Number of Subjects	
	P	Z
Withdrawn from Drug	111	73
Because of Symptoms	8	13
Due to inconvenience	28	30
To seek other therapy	50	15
Reason not stated	25	15
Lost to Follow-Up	41	19
Protocol Reasons	33	19
Toxicity	9	9
Endpoint	21	8
Other	3	2
Total	185	111

Slide 14



Slide 15

Number of Available CD4 Counts

Observed Data:

Week:	8	16	32
Available CD4 (PLC)	371	312	251
Available CD4 (ZDV)	375	325	241

Recensored Data:

Week:	8	16	32
Available CD4 (PLC)	371	312	251
Available CD4 (ZDV)	375	321	160

Slide 16

Estimates of Rx Difference in 019

		Week		
Naive		8	16	32
	$\hat{\theta}$	19.0	23.0	18.0
	SE	11.2	12.9	16.0
Method		8	16	32
	$\hat{\theta}$	19.0	24.0	33.0
	SE	11.4	14.5	20.3

Slide 17

- Simulation 1**
- Design: 500 Simulations of a dataset with $n = 150$ patients randomized to two treatments,
 - $(X_{i1}, X_{i2}, \log(D_i)) \sim$ from a trivariate normal,
 - The correlation of X_{i1} & X_{i2} with $\log(D_i) = 0.70$,
 - Measurements at $(t_1 = 16, t_2 = 48)$,
 - Treatment Difference $(\theta_1 = 25, \theta_2 = 50)$

Slide 18

Results from Sim 1

Method	Week	Average	SD
		of $\hat{\theta}$	of $\hat{\theta}$
Naive	16	22.2	18.5
	48	17.9	27.2
Method	16	26.0	28.4
	48	49.8	30.4

Slide 19

Simulation 2

- How much efficiency is lost if censoring is actually independent?
- Design identical to Simulation 1, except D_i is independent of (X_{i1}, X_{i2})
- 3,000 Datasets were simulated; the point estimates were obtained by both the naive and proposed method.

Slide 20

Results from Sim2

Method	Week	Average of $\hat{\theta}$	SD of $\hat{\theta}$
Naive			
	16	24.9	18.5
	48	49.7	20.9
Method			
	16	25.0	19.2
	48	49.8	22.7

Slide 21

Conclusions

- Method is computationally simple and more flexible than many models that have been proposed.
- Important to consider the effect of treatment on the correlation between X_i and D_i .
- Sometimes you may want to treat withdrawal as categorical.
- Hope to provoke more interest in flexible models for dependent censoring.

Slide 22