Accommodating informative dropout and death: a joint modelling approach for longitudinal and semi-competing risks data

Qiuju Li

MRC Biostatistics Unit, Cambridge, UK

qiuju.li@mrc-bsu.cam.ac.uk

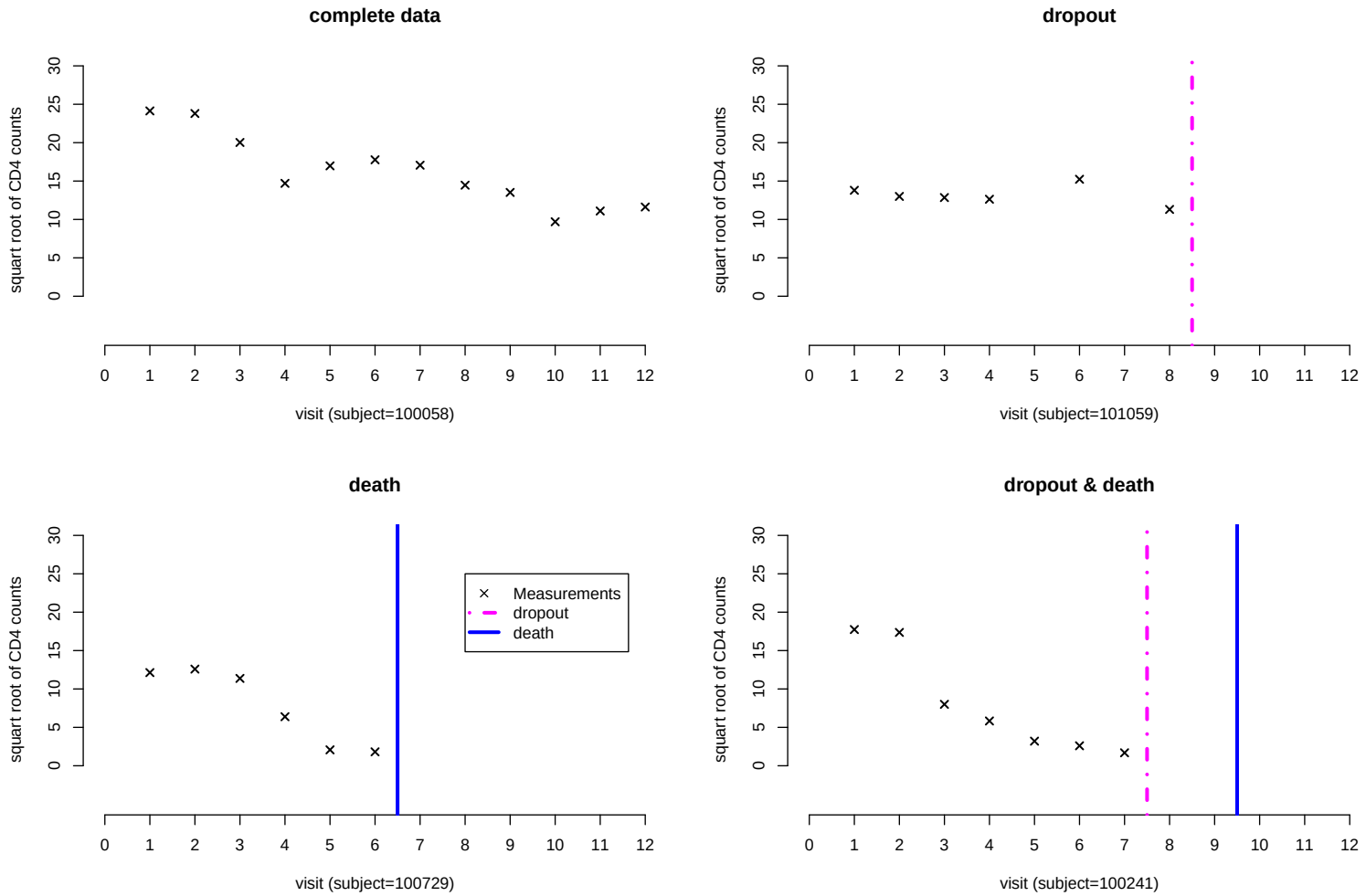
Joint work with Dr. Li Su

Workshop on Flexible Models for Longitudinal and Survival Data with Applications in Biostatistics (Warwick, 27th-29th July, 2015)

Outline

Introduction

- longitudinal and semi-competing risks data, e.g., CD4 counts, dropout and HIV-related death in the HIV epidemiology research study (HERS).



Some concepts

- mortal cohort;
- immortal cohort;
  (Aalen and Gunnes, 2010)

longitudinal profile models:

- unconditional models, e.g., random-effects models $f(Y_i(t))$;
- fully conditional models, e.g., $f(Y_i(t)|S_i = s)$, $s > t$;
- partly conditional models, e,g., $f(Y_i(t)|S_i > t)$ (Kurland and Heagerty, 2005; Kurland *et al.*, 2009);
  - GEE approaches;
  - a likelihood-based joint modelling approach proposed subsequently.

Notation

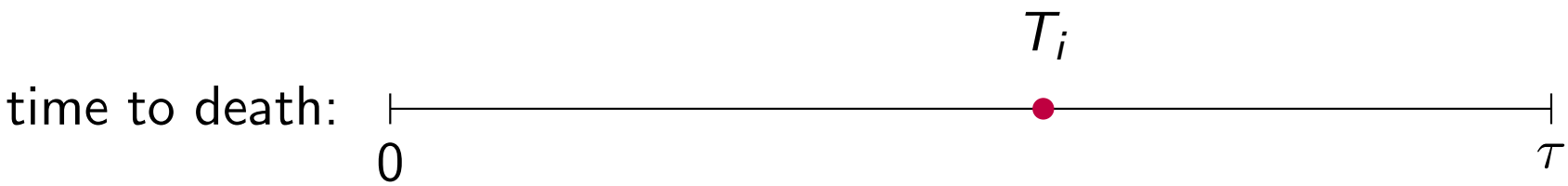
- scheduled repeated measurements of a longitudinal outcome $Y_i = (Y_{i1}, \ldots, Y_{iM})'$, taken at visits $1, \ldots, M$, e.g., $M = 12$ for the HERS data;
- informative dropout and death
  - dropout time denoted by D_i, observed data $D_i^* = min(D_i, S_i, C_i)$, $\delta_i^D = I(D_i \leq S_i, D_i \leq C_i)$;
  - death time denoted by S_i, observed data $S_i^* = min(S_i, C_i)$, $\delta_i^S = I(S_i \leq C_i)$;

  C_i denotes non-informative censoring, e.g., end of study;
- covariates X_i, e.g., sex, treatment arm;

Time-to-event processes

Time-to-event data,

- time to dropout: last visit of follow-up;
- time to death:

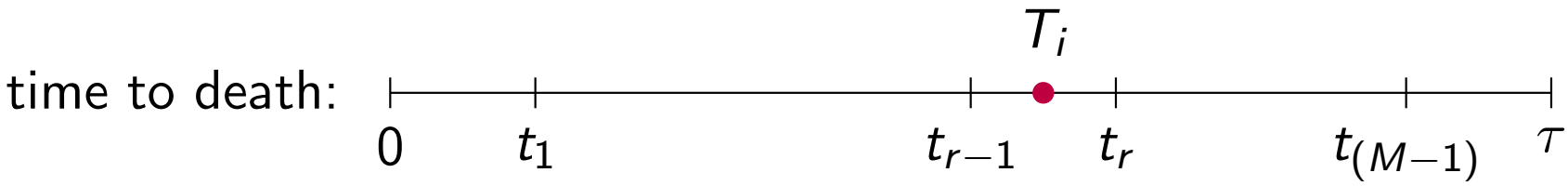


τ: the end of study.

Time-to-event processes

Discrete time-to-event data (*mathematical attractiveness*),

- time to dropout: last visit of follow-up;
- time to death:



τ: the end of study.

the discrete death time $S_i = r$ *(Barrett et al, 2015, JRSSB)*.

Joint models

Joint models for the longitudinal and semi-competing risks data,

$$\begin{cases} Y_{ij} = x_{ij}^T \beta + z_{ij}^T b_i + \epsilon_{ij} \\ Pr(D_i = r | D_i \geq r) = 1 - \Phi(x_{D,ir}^T \alpha_D + \gamma_{D,r}^T W_{D,ir} b_i) \\ Pr(S_i = r | S_i \geq r) = 1 - \Phi(x_{S,ir}^T \alpha_S + \gamma_{S,r}^T W_{S,ir} b_i) \end{cases},$$

- $\Phi(\cdot)$ standard normal cdf;
- β, α_D, α_S regression coefficients; $\gamma_{D,r}$, $\gamma_{S,r}$ association effects;
- random effects $b_i \sim N(0, \Sigma_b)$;
- $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$;
- $W_{D,ir} b_i$, $W_{S,ir} b_i$ vectors of linear combinations of random effects, e.g., $W_{D,ir} b_i = (b_{i0}, b_{i1})^T$;
- conditional independence assumption given random effects b_i;

Likelihood function,

$$\prod_i L_i(\theta; y_i, D_i^*, \delta_i^D, S_i^*, \delta_i^S)$$

$$= \prod_i \int_{-\infty}^{\infty} f(\text{longitudinal data}|\theta, b_i) \times$$

$$Pr(\text{dropout data}|\theta, b_i) \times Pr(\text{death data}|\theta, b_i) \times f(b_i|\theta) db_i.$$

Semi-competing risks data

Four possible scenarios of the observed time-to-event data,

(1) neither dropout nor death:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (0, 0);$$

(2) dropout only:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (1, 0);$$

(3) death only:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (0, 1);$$

(4) both dropout and death:
$$D_i^* = d, S_i^* = s, (\delta_i^D, \delta_i^S) = (1, 1);$$

For the scenario (1), the likelihood function of observed data $\{y_i = (y_{i1}, \ldots, y_{in_i})', D_i^* = d, \delta_i^D = 0, S_i^* = s, \delta_i^S = 0\}$,

$$L_i(\theta; y_i, D_i^*, \delta_i^D, S_i^*, \delta_i^S)$$

$$= \int_{-\infty}^{\infty} \phi(y_i; x_i\beta + z_i b_i, \sigma^2 I_{n_i}) \prod_{k=1}^{d} \Phi(x_{D,ir}^T \alpha_D + \gamma_{D,r}^T W_{D,ir} b_i)$$

$$\prod_{\ell=1}^{s} \Phi(x_{S,ir}^T \alpha_S + \gamma_{S,r}^T W_{S,ir} b_i) \phi(b_i; 0, \Sigma_b) db_i$$

$$= L_{i1}(\cdot \backslash b_i) \Phi^{(d+s)}(A_{ds} + B_{ds} h_i; 0, I_{d+s} + B_{ds} H_i^{-1} B_{ds}^T)$$

- closed-form likelihood (skewed normal distribution, Arnold 2009);
- $L_{i1}(\cdot \backslash b_i)$, h_i, H_i, A_{ds}, B_{ds} function/vectors/matrices free of b_i;
- $\phi(\cdot; \mu, \Sigma)$ and $\Phi^{(d+s)}(\cdot; \mu, \Sigma)$ denote multivariate normal pdf/cdf.

Marginal mean profile conditional on being alive

- Unconditional population mean profile for an immortal cohort

$$E(Y_{ij}|x_{ij}) = x_{ij}^T \beta;$$

- Conditional mean profile given being alive for a mortal cohort, we can compute

$$E(Y_{ij}|x_{ij}, S_i \geq j) = x_{ij}^T \beta + z_{ij}^T E(b_i|S_i \geq j).$$

Analogously, $f(b_i|S_i \geq j)$ is a multivariate skew-normal distribution,

$$f(b_i|S_i \geq j) = f\big(b_i|S_i > (j-1)\big) = \frac{Pr\big(S_j > (j-1)|b_i\big) f(b_i)}{Pr(S_i > (j-1))},$$

closed form of its expectation can be obtained.

Statistical inference

1. Maximum likelihood-based approach (*exact likelihood*);
   - R software utilising *nlminb* or *optim*.
2. Bayesian approach;
   - implemented using WinBUGS.

HERS data description

HIV epidemiology research study (HERS) (Smith *et al.*, 1997)

# of subjects: 850 (HIV positive at baseline)	
CD4 counts	reviewed every 6 months up to 12 visits

time-to-event data	# of subjects
scenario (1): $(\delta_i^D, \delta_i^S) = (0,0)$	374
scenario (2): $(\delta_i^D, \delta_i^S) = (1,0)$	352
scenario (3): $(\delta_i^D, \delta_i^S) = (0,1)$	23
scenario (4): $(\delta_i^D, \delta_i^S) = (1,1)$	78

Objective: study the role of baseline patient characteristics (i.e., viral load, antiretroviral therapy (art), # of symptoms) on variation in longitudinal CD4 counts.

Models proposed for the HERS data

$$
\begin{cases}
Y_{ij} = \beta_0 + \beta_1 \mathit{visit} + \beta_{2\sim4}\text{viral load} + \beta_5 \mathit{symptoms} + \beta_6 \mathit{art} \\
+\beta_{7\sim9}\text{visit*viral load} + \beta_{10} \mathit{visit} * \mathit{symptoms} + \beta_{11} \mathit{visit} * \mathit{art} + b_{i0} + b_{i1} + \epsilon_{ij} \\
\\
Pr(D_i = r | D_i \geq r) = 1 - \Phi(\alpha_{D,i0} + \alpha_{D,i1\sim3}\text{viral load} + \alpha_{D,i4}\mathit{symptoms} \\
\qquad +\alpha_{D,i5}\mathit{art} + \alpha_{D,i6} r + \alpha_{D,i7} r^2 + \gamma_{D,0} b_{i0} + \gamma_{D,1} b_{i1}) \\
\\
Pr(S_i = r | S_i \geq r) = 1 - \Phi(\alpha_{S,i0} + \alpha_{S,i1\sim3}\text{viral load} + \alpha_{S,i4}\mathit{symptoms} \\
\qquad +\alpha_{S,i5}\mathit{art} + \alpha_{S,i6} r + \alpha_{S,i7} r^2 + \gamma_{S,0} b_{i0} + \gamma_{S,1} b_{i1})
\end{cases}
$$

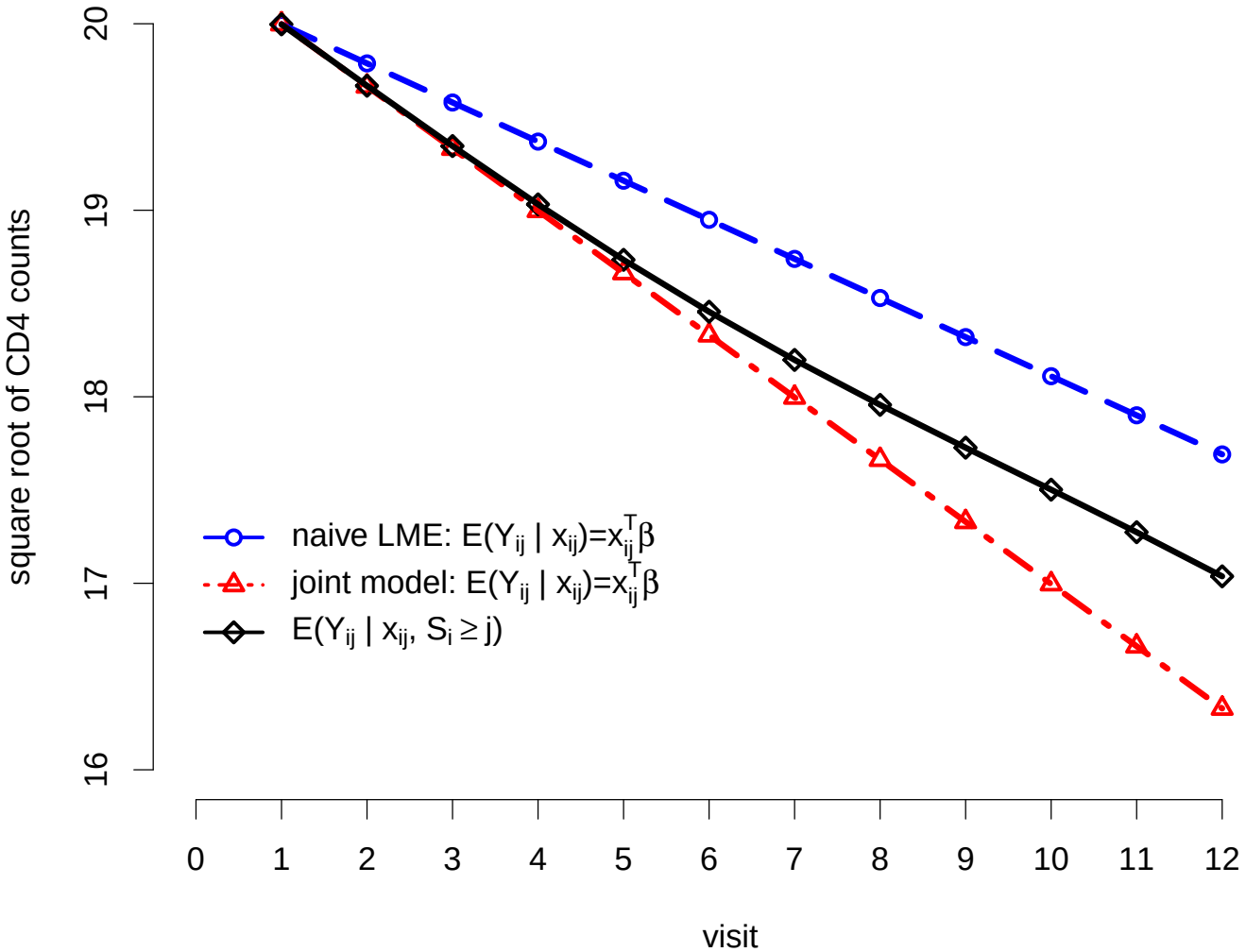
WinBUGS results

Longitudinal process (CD4 counts)

	Joint modelling (WinBUGS)				Linear mixed effects (LME)	
	mean	sd	2.5%	97.5%	estimate	sd
intercept	15.11	0.71	13.76	16.51	14.59	0.70
visit	-0.87	0.13	-1.11	-0.62	-0.57	0.12
viral load [1] (0-500)	10.00	0.79	8.42	11.56	10.52	0.79
viral load [2] (500-5k)	6.61	0.72	5.14	7.97	6.98	0.74
viral load [3] (5k-30k)	2.94	0.82	1.28	4.54	3.21	0.81
symptoms	-0.12	0.20	-0.51	0.28	-0.14	0.21
art at baseline	-4.66	0.43	-5.51	-3.83	-4.76	0.43
visit*viral load [1]	0.47	0.14	0.21	0.74	0.23	0.13
visit*viral load [2]	0.44	0.13	0.19	0.69	0.22	0.12
visit*viral load [3]	0.28	0.14	0.02	0.55	0.15	0.13
visit*symptoms	-0.05	0.03	-0.11	0.01	-0.03	0.03
visit*art	0.12	0.06	-0.01	0.23	0.16	0.06
$\gamma_{D,0}$	0.03	0.01	0.02	0.04	-	-
$\gamma_{D,1}$	0.44	0.05	0.35	0.54	-	-
$\gamma_{S,0}$	0.13	0.02	0.10	0.16	-	-
$\gamma_{S,1}$	1.23	0.17	0.93	1.58	-	-

Marginal mean profiles

Subjects: viral load [1], # of symptoms=1, antiretroviral therapy (art) at baseline;



Conclusions

1. a likelihood-based approach to capture the partly conditional mean profiles, accommodating both informative dropout and death;
2. offer inference for both mortal and immortal cohort;
3. a new model for semi-competing risks data in the joint modelling framework;
4. approach demonstrated by an analysis of the HERS data.

Key references

1. Aalen, Odd O., and Nina Gunnes. "A dynamic approach for reconstructing missing longitudinal data using the linear increments model." Biostatistics 11, no. 3 (2010): 453-472.
2. Arnold, Barry C. "Flexible univariate and multivariate models based on hidden truncation." *Journal of Statistical Planning and Inference* 139, no. 11 (2009): 3741-3749.
3. Barrett, Jessica, Peter Diggle, Robin Henderson and David TaylorRobinson. "Joint modelling of repeated measurements and timetoevent outcomes: flexible model specification and exact likelihood inference." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77, no. 1 (2015): 131-148.
4. Kurland, Brenda F., and Patrick J. Heagerty. "Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths." *Biostatistics* 6, no. 2 (2005): 241-258.

Thank You!

Authors:
Qiuju Li & Li Su
MRC Biostatistics Unit, Cambridge, UK