Model Selection in Bayesian Survival Analysis for a Multi-country Cluster Randomized Trial

Jin Kyung Park

International Vaccine Institute

Min Woo Chae

Seoul National University

R. Leon Ochiai

International Vaccine Institute

Yongdai Kim

Seoul National University

Outlines

1. Motivational data
2. Frailty model, parameters, and priors
3. MCMC algorithm
4. Deviance Information Criterion
5. Simulation
6. Example
7. Summary

1. **Motivational data**

- To assess typhoid vaccine effectiveness, the International Vaccine Institute (IVI) coordinated a clinical trial such that
  - Clustered randomization trial (CRT)
  - Two-year surveillance at Kolkata, India from 2004 to 2006
  - Two-year surveillance at Karachi, Pakistan from 2002/2003 to 2004/2005
  - Time-to-first event of typhoid or censoring
  - Risk factors: gender, age group etc.

Figure 1: Geographic clusters in Kolkata, India

Source: The New England Journal of Medicine, 2009

Table 1: Basic statistics

Country	Clusters no.	Subjects no.	Males no.(%)	Children[†] no.(%)	Cases no.(IR[‡]/1000)
Kolkata, India	80	11958	6157 (51.49)	2192 (18.33)	109 (9.12)
Karachi, Pakistan	120	27231	14090 (51.74)	6481 (23.80)	79 (2.9)

[§] Vaccinees (2 - 16 years aged people)

[†] $\leq$ 5 yrs aged

[‡] Incidence rate(no. of cases/no. of subjects)

- Clustering effect (random effects)
- Variance component of random effects by country
- Common effect of the treatment (vaccine effectiveness)
- We propose a way of selecting a model for the multi-country CRT based on the deviance information criterion (DIC)

2. **Frailty model**

- Observations:

  $(T_{ijk}, \delta_{ijk}, z_{ijk}),\ i = 1, \cdots, n_s,\ j = 1, 2, \cdots, n_{ic},\ k = 1, 2, \cdots, n_{ijl}$

  where

  $T_{ijk} = \min\{X_{ijk}, C_{ijk}\}, \delta_{ijk} = I(X_{ijk} \leq C_{ijk}), z_{ijk} = (z_{ijk1}, \ldots, z_{ijkp})^{'}$,

  X_{ijk} and C_{ijk} are survival and censoring times respectively

  n_s is the number of sites (i.e., countries)

  n_{ic} is the number of clusters of ith site

  n_{ijl} the number of subjects in the jth cluster of ith site

  p is the number of covariates

- Model

$$a_i(t|z_{ijk}, \eta_{ij}) = \eta_{ij} \exp(z_{ijk}^{'}\beta) a_i(t)$$

where $\beta = (\beta_1, \ldots, \beta_p)^{'}$ are the regression coefficients, $a_i(t)$ are baseline hazard functions of ith site and
η_{ij} $(i = 1, \cdots, n_s,\ j = 1, 2, \cdots, n_{ic})$ are cluster specific frailty terms.

- Frailty

$$\eta_{ij} \sim \Gamma(\alpha_i, \alpha_i)$$

It is assumed that η_{ij} are independent random variables distributed by the gamma distribution with mean 1 and variance $1/\alpha_i$. Here, α_i can be considered as the degree of cluster heterogeneity of ith site (Clayton, 1978).

- Parameters: β, $a_i(\cdot)$, α_i
- Priors
  - Normal prior for $\beta_m \sim N(0, \sigma_m^2), \quad m = 1, \ldots, p$
  - Gamma prior for $\alpha_i, \quad \alpha_i \sim \Gamma(\tau_{i1}, \tau_{i2})$
  - Bayesian bootstrap prior for $a_i(\cdot)$

Bayesian Bootstrap (Kim and Lee, 2003)

- To approximate the full Bayesian posterior by Bayesian bootstrap posterior that is proportional to the product of the empirical likelihood and prior
- The conditional likelihood of $\theta = (\beta, a(\cdot))$ given $\eta = (\eta_1, \cdots, \eta_{nc})$

$$
\begin{aligned}
L(\theta|\eta) &= \prod_{i=1}^{n_c}\prod_{j=1}^{n_i}\left(\eta_i \exp(z_{ij}^{'}\beta)a(t_{ij})\right)^{\delta_{ij}} \exp\left(-\int_0^{t_{ij}} \eta_i \exp(z_{ij}^{'}\beta)a(s)ds\right) \\
&= \prod_{i=1}^{n_c}\prod_{j=1}^{n_i}\left(\eta_i \exp(z_{ij}^{'}\beta)dA(t_{ij})\right)^{\delta_{ij}} \exp\left(-\int_0^{t_{ij}} \eta_i \exp(z_{ij}^{'}\beta)dA(s)\right)
\end{aligned}
$$

where $A(t) = \int_0^t a(s)ds$ is the cumulative baseline hazard function.

- Let q be the number of distinct uncensored observations
- Let $0 < t_1 < \cdots < t_q$ be the corresponding ordered uncensored observations
- Then, the empirical likelihood is obtained by assuming that A is a step function having jumps only at $t_1, \ldots, t_q$ and replacing $dA(t)$ by $\Delta A(t) = A(t) - A(t-)$

$$L^E(\theta|\eta) = \prod_{i=1}^{n_c} \prod_{j=1}^{n_i} \left(\eta_i \exp(z_{ij}^{'}\beta) \Delta A(t_{ij}) \right)^{\delta_{ij}} \exp\left(- \sum_{k: t_k \le t_{ij}} \eta_i \exp(z_{ij}^{'}\beta) \Delta A(t_k) \right)$$

- Bayesian bootstrap prior for ΔA (Kim and Lee, 2003)

$$\pi(\Delta A) \propto \prod_{k=1}^{q} \frac{1}{\Delta A(t_k)}$$

- Posterior $\propto$ Empirical likelihood $\times$ BB prior

Advantages of the Bayesian bootstrap

- Finitely many parameters: standard Bayes theorem can be used for posterior calculation

  Parameters: $\beta = (\beta_1, \cdots, \beta_p)'$, $\{\Delta A(t_k), k = 1, \ldots, q\}$, α

- Computation of posterior is easy
- The resulting posterior is
  - always proper,
  - approximates the full Bayesian posterior well, and
  - has desirable large sample properties.
  - Kim and Lee (2003) *Annals of Statistics*

3. MCMC algorithm

- Step 1 : updating cumulative baseline hazards (A_i)
- Step 2 : updating regression coefficients ($\beta_1, \ldots, \beta_p$)
- Step 3 : updating cluster specific random effects (η_{ij})
- Step 4 : updating variance of cluster specific random effects (α_i)

- $\Delta A_i(t_l)|\text{others} \sim \text{Gamma}(a_{il}, b_{il})$, where

$$a_{il} = \sum_{(i,j,k) \in D(t_l)} \delta_{ijk}$$

and

$$b_{il} = \sum_{(i,j,k) \in R(t_l)} \eta_{ij} \exp(z'_{ijk}\beta)$$

where
$R(t) = \{(i, j, k) : T_{ijk} \geq t,\ i = 1, \ldots, n_s,\ j = 1, 2, \cdots, n_c, k = 1, 2, \cdots, n_j\}$
and $D(t) = \{(i, j, k) : T_{ijk} = t, \delta_{ijk} = 1,\ i = 1, \ldots, n_c,\ j = 1, 2, \cdots, n_i, k = 1, 2, \cdots, n_j\ \}$.

- $\pi^p(\beta_1, \ldots, \beta_p|\text{others})$

$$\propto \prod_{l=1}^{q} \left[\prod_{(i,j,k) \in D(t_l)} \eta_{ij} \exp(z'_{ijk}\beta) \right] \exp\left(-\Delta A_i(t_l) \sum_{(i,j,k) \in R(t_l)} \eta_{ij} \exp(z'_{ijk}\beta) \right) \pi(\beta)$$

Metropolis-Hastings algorithm is used for generating β.

- $\eta_{ij}|\text{others} \sim \text{Gamma}(c_{ij}, d_{ij})$, where

$$c_{ij} = \sum_{j=1}^{n_c} \delta_{ijk} + \alpha_i$$

and

$$d_{ij} = \sum_{j=1}^{n_c} \exp(z'_{ijk}\beta)A(t_{ijk}) + \alpha_i.$$

- $\pi^p(\alpha_i|\text{others})$

$$\propto \left[\prod_{j=1}^{n_c} \frac{\alpha_i^{\alpha_i}}{\Gamma(\alpha_i)} \eta_{ij}^{\alpha_i - 1} \exp(-\alpha_i \eta_{ij})\right] \alpha_i^{\tau_{i1}-1} \exp(-\tau_{i2}\alpha_i).$$

Metropolis-Hastings algorithm is used for generating α_i.

Exemplary data: Model selection using CRT data of two countries

Table 2: Models: '-' mean that the corresponding parameters are all different

Model	Treatment effect	Effect of other risk factors	Cluster heterogeneity
M1	-	-	-
M2	$\beta_{11} = \cdots = \beta_{n_c 1}$	-	-
M3	-	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	-
M4	$\beta_{11} = \cdots = \beta_{n_c 1}$	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	-
M5	-	-	$\alpha_1 = \cdots = \alpha_{n_c}$
M6	$\beta_{11} = \cdots = \beta_{n_c 1}$	-	$\alpha_1 = \cdots = \alpha_{n_c}$
M7	-	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	$\alpha_1 = \cdots = \alpha_{n_c}$
M8	$\beta_{11} = \cdots = \beta_{n_c 1}$	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	$\alpha_1 = \cdots = \alpha_{n_c}$

4. **Deviance Information Criterion (DIC)**

- Bayesian Model selection criterion (Spiegelhalter *et al.*, 2002)

$$DIC(\theta) = -4\mathrm{E}_{\theta}[\log f(y|\theta)|y] + 2\log f(y|\tilde{\theta}(y))$$

  where $f(y|\theta)$ be the density of data y and parameter of interest, θ

- DIC is defined as the sum of the expected deviance and the effective number of parameters

- A simple and intuitively appealing extension of the Akaike Information Criterion (AIC, Akaike 1973)

Extensions of DIC (Celux *et al.*, 2006)

- Observed DIC
  - Use the marginal likelihood to define the DIC
  - Often the observed DIC is difficult to calculate

- Alternatives
  - Complete DIC
  - Conditional DIC

Complete DIC

- Defined as the posterior expectation of the DIC obtained assuming the random effects are known

$$\mathrm{DIC}_{\mathrm{comp1}} = -4\mathrm{E}_{\theta,w}[\log f(y,w|\theta)|y] + 2\mathrm{E}_w[\log f(y,w|\tilde{\theta}(y,w))|y]$$

  where

  - $f(y,w|\theta)$ be the density of y and w
  - y and w represent observed data and random effects, respectively

- Computationally demanding when $\tilde{\theta}(y,w)$ does not have a closed form solution

Complete DIC type 2

- Replace the term $\mathrm{E}_w[\log f(y,w|\tilde{\theta}(y,w))|y]$ by $\log f(y,\tilde{w}(y)|\tilde{\theta}(y))$

$$\begin{aligned}\mathrm{DIC}_{\mathrm{comp1}} &= -4\mathrm{E}_{\theta,w}[\log f(y,w|\theta)|y] + 2\mathrm{E}_w[\log f(y,w|\tilde{\theta}(y,w))|y] \\ \mathrm{DIC}_{\mathrm{comp2}} &= -4\mathrm{E}_{\theta,w}[\log f(y,w|\theta)|y] + 2\log f(y,\tilde{w}(y)|\tilde{\theta}(y))\end{aligned}$$

- $(\tilde{w}(y),\tilde{\theta}(y))$ is an reasonable estimator of (w,θ) by treating w as an additional parameter (of interest)

- Tends to overestimate the effective number of parameters

Complete DIC type 3

- Replace $\tilde{\theta}(y, w)$ in the $\mathrm{DIC}_{\mathrm{comp1}}$ by $\tilde{\theta}(y)$

$$\begin{aligned}
\mathrm{DIC}_{\mathrm{comp1}} &= -4\mathrm{E}_{\theta,w}[\log f(y, w|\theta)|y] + 2\mathrm{E}_w[\log f(y, w|\tilde{\theta}(y, w))|y] \\
\mathrm{DIC}_{\mathrm{comp2}} &= -4\mathrm{E}_{\theta,w}[\log f(y, w|\theta)|y] + 2\log f(y, \tilde{w}(y)|\tilde{\theta}(y)) \\
\mathrm{DIC}_{\mathrm{comp3}} &= -4\mathrm{E}_{\theta,w}[\log f(y, w|\theta)|y] + 2\mathrm{E}_w[\log f(y, w|\tilde{\theta}(y))|y]
\end{aligned}$$

- If we let $\tilde{\theta} = \mathrm{E}_\theta[\theta|y]$, which is easily obtained using MCMC samples

Conditional DIC

- The conditional likelihood of y given w and θ is used

$$\mathrm{DIC}_{\mathrm{cond1}} = -4\mathrm{E}_{\theta,w}[\log f(y|w,\theta)|y] + 2\mathrm{E}_{w}[\log f(y|w,\tilde{\theta}(y,w))|y]$$

$$\mathrm{DIC}_{\mathrm{cond2}} = -4\mathrm{E}_{\theta,w}[\log f(y|w,\theta)|y] + 2\log f(y|\tilde{w}(y),\tilde{\theta}(y))$$

$$\mathrm{DIC}_{\mathrm{cond3}} = -4\mathrm{E}_{\theta,w}[\log f(y|w,\theta)|y] + 2\mathrm{E}_{w}[\log f(y|w,\tilde{\theta}(y))|y]$$

Note that the $\mathrm{DIC}_{\mathrm{cond1}}, \mathrm{DIC}_{\mathrm{cond2}}$ and $\mathrm{DIC}_{\mathrm{cond3}}$ are conditional likelihood versions of the $\mathrm{DIC}_{\mathrm{comp1}}, \mathrm{DIC}_{\mathrm{comp2}}$ and $\mathrm{DIC}_{\mathrm{comp3}}$, respectively.

5. **Simulations**

- Complete/Conditional DIC type 2 and 3 are used for model selection
- Generate 100 data sets with
  - no. of sites, $n_s = 2$
  - no. of clusters per site, $n_c = 20$
  - no. of subjects per cluster, $n_k = 5$
- Generate Z_1 and Z_2 as covariates independently from the symmetric Bernoulli distribution with success probability 0.5.
- For the baseline hazard functions, we let $a_1(t) = 1$ and $a_2(t) = 2$

- Consider the four models:

  1. Equal parameter of interest and equal nuisance parameters

     SM1: $\beta_{11} = \beta_{21} = -1.0, \beta_{12} = \beta_{22} = 0.5, \alpha_1 = \alpha_2 = 0.5$

  2. Different parameter of interest and equal nuisance parameters

     SM2: $(\beta_{11}, \beta_{21}) = (-1.0, 0.0)$ and $\beta_{12} = \beta_{22} = 0.5, \alpha_1 = \alpha_2 = 0.5$

  3. Equal parameter of interest and different nuisance parameters

     SM3: $\beta_{11} = \beta_{21} = -1.0, (\beta_{12}, \beta_{22}) = (1.0, 0.0), (\alpha_1, \alpha_2) = (0.3, 0.7)$

  4. Different parameter of interest and different nuisance parameters

     SM4: $(\beta_{11}, \beta_{21}) = (-1.0, 0.0), (\beta_{12}, \beta_{22}) = (1.0, 0.0), (\alpha_1, \alpha_2) = (0.3, 0.7)$

Table 3: No frailty: the frequencies of the models selected by DICs among 100 data sets

Censoring prob.	Simulation Model	Partial likelihood[§]				Full likelihood			
		SM1	SM2	SM3	SM4	SM1	SM2	SM3	SM4
0.0	SM1	100	0	0	0	99	0	1	0
	SM2	21	79	0	0	37	63	0	0
	SM3	25	0	75	0	15	0	85	0
	SM4	2	11	19	68	6	17	20	57
0.5	SM1	94	4	2	0	89	7	3	1
	SM2	22	78	0	0	20	77	1	2
	SM3	24	0	75	1	24	2	69	5
	SM4	5	15	20	60	7	24	18	51

[§] *Selection by using AIC*

Table 4: Frailty: the frequencies of the models selected by COMPLETE DICs among 100 data sets

True model		COMPLETE DICs							
Censoring prob.	Simulation Model	Type 2				Type 3			
		SM1	SM2	SM3	SM4	SM1	SM2	SM3	SM4
0.0	SM1	30	34	10	26	42	34	6	18
	SM2	15	49	13	23	19	53	12	16
	SM3	17	27	24	32	19	23	34	24
	SM4	13	23	16	48	12	22	16	50
0.5	SM1	16	37	12	35	21	35	11	33
	SM2	23	26	19	32	24	29	17	30
	SM3	15	19	19	47	18	17	23	42
	SM4	9	19	23	49	9	19	24	48

Table 5: Frailty: the frequencies of the models selected by CONDITIONAL DICs among 100 data sets

True model		CONDITIONAL DICs							
Censoring prob.	Simulation Model	Type 2				Type 3			
		SM1	SM2	SM3	SM4	SM1	SM2	SM3	SM4
0.0	SM1	100	0	0	0	100	0	0	0
	SM2	23	77	0	0	24	76	0	0
	SM3	14	0	86	0	20	0	80	0
	SM4	6	17	20	57	7	18	19	56
0.5	SM1	89	7	3	1	89	7	3	1
	SM2	20	77	1	2	20	77	1	2
	SM3	24	2	69	5	25	3	68	4
	SM4	7	23	18	52	7	24	19	50

Summary of simulations

- DIC using full likelihood selects model well where the data has no frailties
- Complete DICs highly depend on frailty distribution
- Conditional DICs is proposed to be used for the model selection of the frailty models

$$DIC_{cond2} = -4E_{\theta,w}[\log f(y|w,\theta)|y] + 2\log f(y|\tilde{w}(y),\tilde{\theta}(y))$$

$$DIC_{cond3} = -4E_{\theta,w}[\log f(y|w,\theta)|y] + 2E_w[\log f(y|w,\tilde{\theta}(y))|y]$$

$$\begin{aligned} DIC_{comp2} &= -4E_{\theta,w}[\log f(y|w,\theta)|y] + 2\log f(y|\tilde{w}(y),\tilde{\theta}(y)) \\ &\quad -4E_{\theta,w}[\log \pi(w|\theta)] + 2\log \pi(\tilde{w}(y)|\tilde{\theta}(y)) \end{aligned}$$

$$\begin{aligned} DIC_{comp3} &= -4E_{\theta,w}[\log f(y|w,\theta)|y] + 2E_w[\log f(y|w,\tilde{\theta}(y))|y] \\ &\quad -4E_{\theta,w}[\log \pi(w|\theta)] + 2E_w[\log \pi(w|\tilde{\theta}(y))] \end{aligned}$$

where $w = \eta_i$ and the distribution of η_i follows $\eta_i \sim Gamma(\alpha,\alpha)$

6. **Exemplary data: model selection using CRT data of two countries**

Table 6: Models: '-' mean that the corresponding parameters are all different

Model	Treatment effect	Effect of other risk factors	Cluster heterogeneity
M1	-	-	-
M2	$\beta_{11} = \cdots = \beta_{n_c 1}$	-	-
M3	-	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	-
M4	$\beta_{11} = \cdots = \beta_{n_c 1}$	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	-
M5	-	-	$\alpha_1 = \cdots = \alpha_{n_c}$
M6	$\beta_{11} = \cdots = \beta_{n_c 1}$	-	$\alpha_1 = \cdots = \alpha_{n_c}$
M7	-	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	$\alpha_1 = \cdots = \alpha_{n_c}$
M8	$\beta_{11} = \cdots = \beta_{n_c 1}$	$\beta_{1(-1)} = \cdots = \beta_{n_c(-1)}$	$\alpha_1 = \cdots = \alpha_{n_c}$

MCMC sampling

- 500 times of burn-in
- no. of repetitions = 10000
- no. of updates = 1000

Table 7: Results of fitting data to frailty model: Type 2 DICs

	Conditional DIC		Complete DIC	
Model	DIC	p_D	DIC	p_D
M1	3237.64	63.55	3572.15	29.78
M2	3236.45	63.39	3568.77	27.85
M3	3237.00	61.98	3573.56	29.85
M4	3235.55	61.53	3570.57	28.05
M5	3244.01	61.27	3651.15	75.13
M6	3241.49	61.26	3644.61	71.22
M7	3243.50	59.46	3651.79	75.09
M8	3240.88	59.29	3645.45	70.71

Table 8: Results of fitting data to frailty model: Type 3 DICs

	Conditional DIC		Complete DIC	
Model	DIC	p_D	DIC	p_D
M1	3177.84	3.75	3536.01	-6.35
M2	3176.70	3.64	3533.86	-7.05
M3	3177.35	2.32	3535.63	-8.07
M4	3176.11	2.09	3534.09	-8.42
M5	3186.64	3.90	3574.41	-1.60
M6	3183.93	3.70	3571.28	-2.11
M7	3186.50	2.47	3573.49	-3.21
M8	3183.61	2.25	3571.40	-3.33

Table 9: Results of fitting data

Model	Country	Effect	coeff	HR†	95% CI of HR	
M4	I+II	Vaccine	-0.754	0.470	(0.306,	0.707)
		Child	0.677	1.968	(1.446,	2.670)
		Male	0.046	1.047	(0.387,	1.454)
Variance of frailties of		Kolkata	$1/\hat{\alpha_1}^B = 0.671$		(0.387,	1.236)
		Karachi	$1/\hat{\alpha_2}^B = 1.356$		(0.768,	2.560)
M2	I+II	Vaccine	-0.754	0.470	(0.310,	0.728)
	I	Child	0.698	2.010	(1.287,	3.007)
		Male	0.166	1.181	(0.811,	1.740)
	II	Child	0.701	2.016	(1.262,	3.076)
		Male	-0.081	0.923	(0.608,	1.402)
Variance of frailties of		Kolkata	$1/\hat{\alpha_1}^B = 0.686$		(0.403,	1.233)
		Karachi	$1/\hat{\alpha_2}^B = 1.360$		(0.772,	2.756)

$\hat{\alpha_i}^B$ is the Bayes estimate of $\alpha_i (i = 1, 2)$

Summary of the example

- Model selection is done using the conditional DICs
- Data of CRT from two countries are combined
- The cluster effects are heterogeneous among countries (variance component of countries)

7. **Summary**

- Introduce frailty model for the cluster randomized trial
  - Bayesian bootstrap prior for baseline hazard function
  - Gamma prior frailty
- Introduce model selection for frailty model
  - Complete DIC
  - Conditional DIC
- Perform intensive simulation for the proportional hazards model and frailty model
- Conditional DIC is proposed for the frailty model selection

Acknowledgement

This work was supported by the Bill and Melinda Gates Foundation through the Diseases of the Most Impoverished (DOMI) program, administered by the International Vaccine Institute (IVI), Seoul, Korea.

INTERNATIONAL VACCINE INSTITUTE

Vaccine, Children, and a Better World