

# **Modeling the clustered interval-censored failure time data with informative cluster size**

수원대학교 통계정보학과

김 진 흠

2013. 10. 19

# Informative cluster size

- 같은 군집 안에 있는 개체들의 생존시간이 군집의 크기와 연관되어 있을 때
- 몇 가지 예
  - 독성학 연구: 살충제가 새끼들의 생존시간에 미치는 연구에서 새끼의 수가 작을수록 생존시간이 짧음(군집: 같은 어미의 새끼)
  - 치아 연구: 어금니가 많을수록 어금니의 생존시간이 길어짐(군집: 어금니) [fig.pdf](#)

# MM(marginal model) 방법

- Leave the dependence structure of the correlated failure times unspecified
  - Lee 등(1992)
  - Wei 등(1989)
  - Huster 등(1989)
  - Use corrected(sandwich) standard error
- Informative cluster size
  - Cong 등(2007), Williamson 등(2008): right-censored
  - Zhang & Sun(2011): interval-censored, parametric approach

# WCR(within-cluster resampling) 방법

- GEE 대안
  - 근사적으로는 상관구조의 misspecification에 대해 로버스트 하지만 finite dataset에 대해서는 로버스트 하지 않음 (Mantel & Leroux, 1996)
- Hoffman 등(2001)이 제안
  - 각 군집에서 개체 하나씩만 랜덤 추출
  - resampled dataset 에 있는 개체들은 서로 독립! 따라서 기존 연구방법 적용 가능!
  - resampled data 로부터 반복적으로 모수를 추정한 후 평균 값을 최종 추정량으로
  - resampling 으로 인해 많은 시간이 필요하지만 추정량의 성질은 우수(일치성, 점근 정규성 만족)

# 모형

- $T_{ij}$ :  $i$ 번째 군집 안에 있는  $j$ 번째 개체의 생존시간 ( $i = 1, \dots, n; j = 1, \dots, n_i$ )
- Lee 등(1992)의 모형(MM 모형)

$$\lambda(t|x_{ij}) = \lambda_0(t)\exp(\beta'x_{ij})$$

- $T_{ij}$ 는 관측되지 않고 대신  $(L_{ij}, R_{ij}]$ 이 관측됨
- 관측된 자료는  $\{(L_{ij}, R_{ij}], x_{ij}: i = 1, \dots, n; j = 1, \dots, n_i\}$
- 가정:  $x_{ij}$ 가 주어졌을 때  $T_{ij}$ 는  $L_{ij}, R_{ij}$ 와 서로 독립

# WEE(weighted estimating equation) 방법: 우도 함수

- $\omega_i$ :  $i$ 번째 군집에 있는 각 개체가 우도 함수에 미치는 기여도
- $D_i$ :  $i$ 번째 군집에 있는 개체 중에서 구간중도절단 된 개체들의 집합
- $R_i$ :  $i$ 번째 군집에 있는 개체 중에서 우중도절단 된 개체들의 집합
- $i$ 번째 군집의 의사 우도 함수:  $i$ 번째 군집에 있는 개체들의  $T_{ij}$  가 서로 독립이라고 가정하면,

$$L_i(\lambda_0, \beta | x_{ij}) = \prod_{j \in D_i} \{S(L_{ij} | x_{ij}) - S(R_{ij} | x_{ij})\}^{\omega_i} \prod_{j \in R_i} S(L_{ij} | x_{ij})^{\omega_i}$$

- 군집 전체의 의사 우도 함수:

$$\begin{aligned} L_f(\lambda_0, \beta) &= \prod_{i=1}^n L_i(\lambda_0, \beta | x_{ij}) \\ &= \prod_{i=1}^n \left\{ \prod_{j \in D_i} \{ \exp(-\Lambda_0(L_{ij}) \exp(\beta' x_{ij})) \right. \\ &\quad \left. - \exp(-\Lambda_0(R_{ij}) \exp(\beta' x_{ij})) \}^{\omega_i} \times \prod_{j \in R_i} \exp(-\omega_i \Lambda_0(L_{ij}) \exp(\beta' x_{ij})) \right\} \end{aligned}$$

# Re-parameterization

- $0 = s_0 < s_1 < \dots < s_m < s_{m+1} = \infty$ : equivalence class 의 중간값
- $\Lambda_0(s_q) = \sum_{k=0}^q \exp(\alpha_k) \equiv a_q, \quad q = 0, \dots, m+1 \quad \text{단, } \alpha_0 = -\infty, \alpha_{m+1} = \infty$
- $\alpha_{ijq} = I(s_q \in (L_{ij}, R_{ij}])$ ,  $q = 1, \dots, m$
- 군집 전체의 로그 의사 우도 함수:  $\alpha = (\alpha_1, \dots, \alpha_m)'$  로 놓으면,

$$\begin{aligned}
 l_f(\alpha, \beta) &= \log L_f(\lambda_0, \beta) \\
 &= \sum_{i=1}^n \omega_i \left\{ \sum_{j \in D_i} \log \left\{ \sum_{q=1}^{m+1} \alpha_{ijq} \left\{ \exp(-a_{q-1} \exp(\beta' x_{ij})) - \exp(-a_q \exp(\beta' x_{ij})) \right\} \right\} \right. \\
 &\quad \left. - \sum_{j \in R_i} \sum_{q=1}^{m+1} I(L_{ij} \in [s_{q-1}, s_q)) a_{q-1} \exp(\beta' x_{ij}) \right\}
 \end{aligned}$$

# Weighted score functions

- 표기를 간편하게 하기 위해

$$f_{ijq} = -a_q \exp(\beta' x_{ij}) \exp(-a_q \exp(\beta' x_{ij})); \quad b_{ijq} = \exp(a_q + \beta' x_{ij}),$$

$$c_{ijq} = \sum_{l=q}^{m+1} (a_{ijl} - a_{ijl+1}) \exp(-a_l \exp(\beta' x_{ij})),$$

$$g_{ij} = \sum_{q=1}^{m+1} \alpha_{ijq} \{ \exp(-a_{q-1} \exp(\beta' x_{ij})) - \exp(-a_q \exp(\beta' x_{ij})) \}$$

- Weighted score functions

$$U_{\beta} = \frac{\partial l_f}{\partial \beta} = \sum_{i=1}^n \omega_i U_{\beta}^i,$$

$$U_{\alpha} = \frac{\partial l_f}{\partial \alpha} = \sum_{i=1}^n \omega_i U_{\alpha}^i, \quad U_{\alpha}^i = (U_{\alpha_1}^i, \dots, U_{\alpha_m}^i)'$$

단,

$$U_{\beta}^i = \sum_{j \in D_i} x_{ij} \frac{\sum_{q=1}^{m+1} \alpha_{ijq} (f_{ijq-1} - f_{ijq})}{g_{ij}} - \sum_{j \in R_i} x_{ij} \sum_{q=1}^{m+1} I(L_{ij} \in [s_{q-1}, s_q]) a_{q-1} \exp(\beta' x_{ij}),$$

$$U_{\alpha_q}^i = \sum_{j \in D_i} \frac{b_{ijq} c_{ijq}}{g_{ij}} - \sum_{j \in R_i} b_{ijq} I(L_{ij} \in [s_q, \infty)), \quad q = 1, \dots, m$$

# Parameter estimation

- Weighted score equations,  $U_\beta = 0, U_\alpha = 0$  를 만족하는 해를  $\hat{\beta}^{wee}, \hat{\alpha}^{wee}$  으로 놓으면 이는 Newton-Raphson 알고리즘을 반복적으로 적용하여 얻을 수 있음
- 따라서 기저위험함수에 대한 추정량은 아래와 같이 얻어짐

$$\hat{\Lambda}_0(s_q) = \sum_{k=1}^q \exp(\hat{\alpha}_k^{wee}), q = 1, \dots, m$$

# Asymptotic properties

- 적절한 조건하에서(Huang & Wellner, 1997)  
 $\sqrt{n}(\hat{\beta}^{wee} - \beta) \xrightarrow{d} N(0, \Sigma)$  이고,  $\Sigma$  에 대한 일치 추정량은  
 아래와 같아 sandwich-type 추정량 임

$$\hat{\Sigma}^{wee}$$

$$= I_{11|2}(\hat{\beta}^{wee}, \hat{\alpha}^{wee})^{-1} V(\hat{\beta}^{wee}, \hat{\alpha}^{wee}) I_{11|2}(\hat{\beta}^{wee}, \hat{\alpha}^{wee})^{-1}$$

$$\text{단, } V = \sum_{i=1}^n \omega_i^2 U_{\beta}^i U_{\beta}^{i'}, \quad I_{11|2} = I_{11} - I_{12} I_{22}^{-1} I_{21},$$

$$I_{11} = -\frac{\partial^2 l_f}{\partial \beta \partial \beta'}, \quad I_{12} = -\frac{\partial^2 l_f}{\partial \alpha \partial \beta'} = I'_{21}, \quad I_{22} = -\frac{\partial^2 l_f}{\partial \alpha \partial \alpha'}$$

# WCR 방법

- 각  $b = 1, \dots, B$  에 대해,  $\left((L_i^b, R_i^b], x_i^b\right)$ 를  $\left\{\left((L_{ij}, R_{ij}], x_{ij}\right): j = 1, \dots, n_i\right\}$  에서 랜덤 추출한 표본으로 놓음
- $\hat{\beta}^{wee^b}, \hat{\alpha}^{wee^b}$ 를  $\left\{\left((L_i^b, R_i^b], x_i^b\right): i = 1, \dots, n\right\}$  로부터 추정된  $\beta, \alpha$  의 최대우도추정량으로 놓음 ( $\omega_i = 1$ )
- WCR 추정량을 다음과 같이 정의

$$\hat{\beta}^{wcr} = \frac{1}{B} \sum_{b=1}^B \hat{\beta}^{wee^b}, \quad \hat{\alpha}^{wcr} = \frac{1}{B} \sum_{b=1}^B \hat{\alpha}^{wee^b}$$

# Asymptotic properties

- 각  $b = 1, \dots, B$  에 대해, 추정량  $\hat{\beta}^{wee^b}$  은 적절한 조건하에서 (Huang & Wellner, 1997)  
$$\sqrt{n}(\hat{\beta}^{wee} - \beta) \xrightarrow{d} N(0, I_{11|2}^{-1})$$
- 따라서  $\hat{\beta}^{wcr}$  도  $\sqrt{n}(\hat{\beta}^{wcr} - \beta) \xrightarrow{d} N(0, \Sigma)$  이고  $\Sigma$ 에 대한 일치 추정량은 아래와 같이 주어짐(Hoffman 등, 2001)

$$\begin{aligned} \hat{\Sigma} = & \frac{1}{B} \sum_{b=1}^B I_{11|2} \left( \hat{\beta}^{wee^b}, \hat{\alpha}^{wee^b} \right)^{-1} \\ & - \frac{1}{B} \sum_{b=1}^B \left( \hat{\beta}^{wee^b} - \hat{\beta}^{wcr} \right) \left( \hat{\beta}^{wee^b} - \hat{\beta}^{wcr} \right)', \end{aligned}$$

# Simulation setup

- $T_{ij}$ : Cox model with a positive stable frailty  $w_i$  즉,

$$\lambda(t|x_{ij}, w_i) = \lambda_0(t)w_i \exp(\beta_1 x_{i1} + \beta_2 x_{ij2})$$

- $w_i$ : positive stable distribution with a parameter of  $\alpha$  which represents the degree of correlation between cluster members
  - $x_{i1}$ :  $B(1,0.5)$ , cluster-specific covariate
  - $x_{ij2}$ :  $U(-0.5,0.5)$ , subject-specific covariate
  - $\beta_1 = \beta_2 = 1, \lambda_0(t) = 0.3$
  - marginal distribution parameters, 즉  $\beta_1^* = \alpha\beta_1, \beta_2^* = \alpha\beta_2, \Lambda_0^*(t) = t^\alpha \Lambda_0(t)$
- informative cluster size( $n_i$ ): 1~6
    - $w_i$ 가 중앙값보다 작으면  $B(7,0.75)$ , 그렇지 않으면  $B(7,0.25)$ 에서 생성
    - 0,7 은 제외

# Simulation setup

- 14 discrete examination time points: 0.1, 0.6, ..., 6.1, 6.6
  - Miss the scheduled examination times with prob. of 0.3 for the first seven time points and of 0.5 for the last seven time points
  - $T_{ij}$  가 속하는 구간을 결정, 즉  $(L_{ij}, R_{ij}]$
  - $T_{ij} \leq 0.1$  이면 left-censored at 0.1;  $T_{ij} \geq 6.6$  이면 right-censored at 6.6
- 가중값:  $\omega_i = 1/n_i$
- 군집의 크기: 150, 300
- 500번 반복,  $B = 1000$

# Simulation results: non-informative case

| $n$ | $\alpha$ | True value | WEE       |       |       |       | UWE       |       |       |       | WCR       |       |       |       |
|-----|----------|------------|-----------|-------|-------|-------|-----------|-------|-------|-------|-----------|-------|-------|-------|
|     |          |            | Bias      | SD    | SEM   | CP    | Bias      | SD    | SEM   | CP    | Bias      | SD    | SEM   | CP    |
| 150 | 0.2      | 0.2        | -1.92E-02 | 0.190 | 0.189 | 0.948 | -1.18E-02 | 0.210 | 0.209 | 0.948 | -1.83E-02 | 0.191 | 0.189 | 0.946 |
|     |          | 0.2        | 7.50E-03  | 0.228 | 0.221 | 0.936 | 7.18E-03  | 0.185 | 0.184 | 0.948 | 8.68E-03  | 0.232 | 0.223 | 0.930 |
|     | 0.5      | 0.5        | 1.55E-02  | 0.158 | 0.160 | 0.948 | 1.21E-02  | 0.168 | 0.168 | 0.934 | 2.01E-02  | 0.160 | 0.160 | 0.946 |
|     |          | 0.5        | -1.88E-02 | 0.214 | 0.203 | 0.944 | -3.38E-03 | 0.178 | 0.170 | 0.942 | -1.40E-02 | 0.219 | 0.207 | 0.926 |
|     | 0.8      | 0.8        | 2.96E-03  | 0.137 | 0.134 | 0.942 | 1.93E-03  | 0.131 | 0.128 | 0.932 | 1.38E-02  | 0.140 | 0.133 | 0.940 |
|     |          | 0.8        | 8.43E-03  | 0.220 | 0.198 | 0.926 | 5.15E-03  | 0.182 | 0.167 | 0.908 | 2.00E-02  | 0.225 | 0.202 | 0.904 |
| 300 | 0.2      | 0.2        | -5.61E-03 | 0.135 | 0.133 | 0.942 | -6.87E-03 | 0.153 | 0.147 | 0.930 | -5.10E-03 | 0.136 | 0.133 | 0.940 |
|     |          | 0.2        | -4.06E-04 | 0.159 | 0.156 | 0.948 | 1.42E-03  | 0.130 | 0.129 | 0.956 | 4.56E-04  | 0.160 | 0.161 | 0.950 |
|     | 0.5      | 0.5        | -5.89E-03 | 0.119 | 0.112 | 0.930 | -5.67E-03 | 0.124 | 0.118 | 0.948 | -3.80E-03 | 0.120 | 0.112 | 0.930 |
|     |          | 0.5        | 8.89E-05  | 0.149 | 0.146 | 0.944 | 1.37E-03  | 0.120 | 0.121 | 0.946 | 2.56E-03  | 0.150 | 0.152 | 0.954 |
|     | 0.8      | 0.8        | 5.23E-03  | 0.097 | 0.094 | 0.956 | 3.12E-03  | 0.095 | 0.090 | 0.936 | 1.04E-02  | 0.098 | 0.095 | 0.952 |
|     |          | 0.8        | 2.61E-03  | 0.146 | 0.141 | 0.938 | 4.55E-03  | 0.122 | 0.118 | 0.926 | 8.21E-03  | 0.148 | 0.148 | 0.930 |

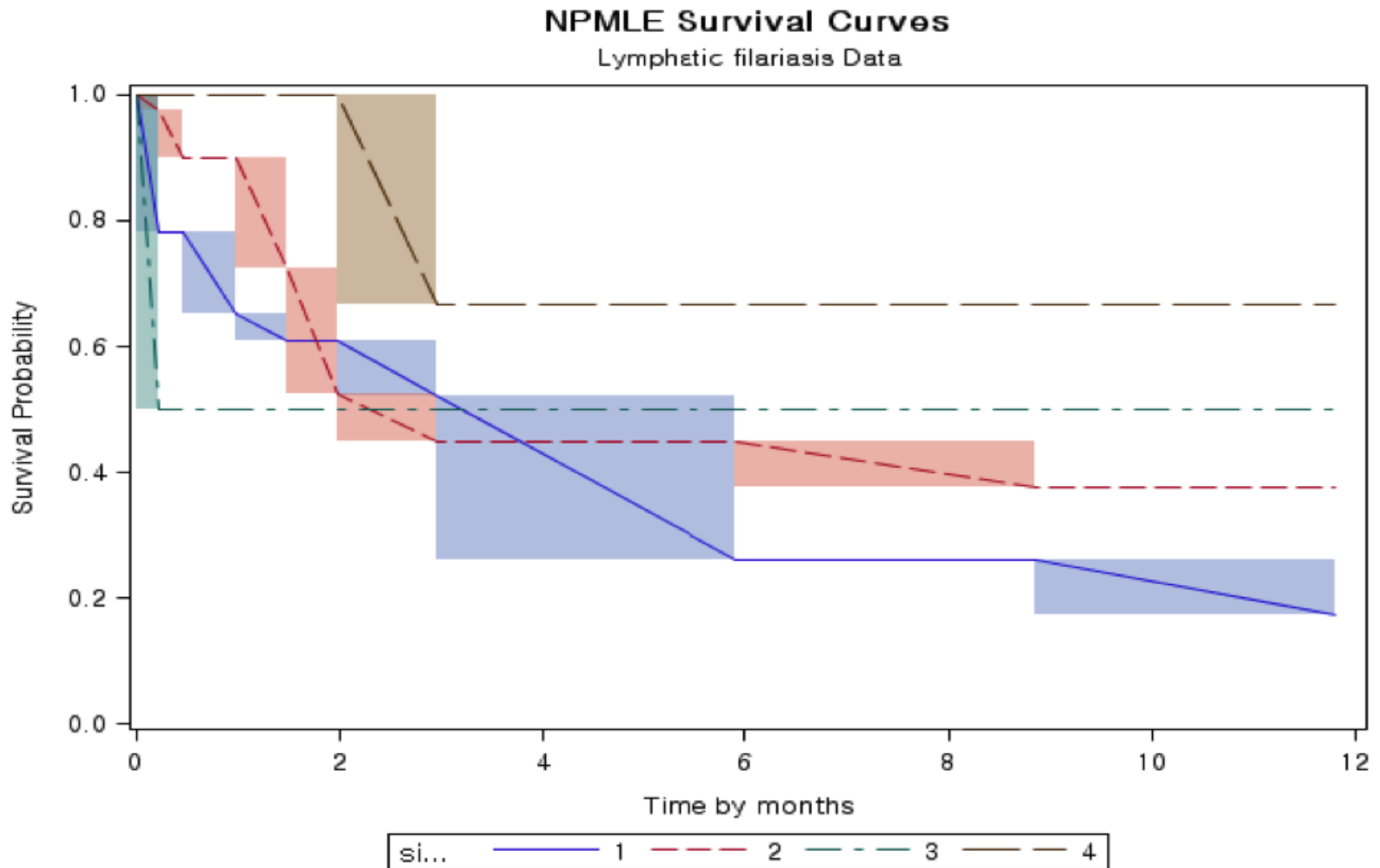
# Simulation results: informative case

| n   | $\alpha$ | True value | WEE       |       |       |       | UWE      |       |       |       | WCR       |       |       |       |
|-----|----------|------------|-----------|-------|-------|-------|----------|-------|-------|-------|-----------|-------|-------|-------|
|     |          |            | Bias      | SD    | SEM   | CP    | Bias     | SD    | SEM   | CP    | Bias      | SD    | SEM   | CP    |
| 150 | 0.2      | 0.2        | -6.87E-03 | 0.184 | 0.187 | 0.952 | 7.10E-02 | 0.213 | 0.214 | 0.944 | -5.76E-03 | 0.185 | 0.187 | 0.952 |
|     |          | 0.2        | -1.63E-02 | 0.199 | 0.200 | 0.952 | 6.89E-02 | 0.196 | 0.199 | 0.940 | -1.45E-02 | 0.202 | 0.203 | 0.946 |
|     | 0.5      | 0.5        | 1.55E-02  | 0.153 | 0.152 | 0.938 | 1.21E-01 | 0.161 | 0.157 | 0.848 | 2.12E-02  | 0.155 | 0.151 | 0.932 |
|     |          | 0.5        | -7.11E-04 | 0.195 | 0.184 | 0.940 | 9.43E-02 | 0.189 | 0.178 | 0.912 | 6.28E-03  | 0.199 | 0.183 | 0.926 |
|     | 0.8      | 0.8        | 8.05E-03  | 0.137 | 0.127 | 0.924 | 6.51E-02 | 0.125 | 0.118 | 0.906 | 2.25E-02  | 0.140 | 0.125 | 0.908 |
|     |          | 0.8        | 1.34E-02  | 0.187 | 0.188 | 0.942 | 7.21E-02 | 0.174 | 0.173 | 0.930 | 2.91E-02  | 0.193 | 0.186 | 0.918 |
| 300 | 0.2      | 0.2        | 4.79E-03  | 0.132 | 0.131 | 0.944 | 8.01E-02 | 0.158 | 0.151 | 0.898 | 5.42E-03  | 0.132 | 0.131 | 0.944 |
|     |          | 0.2        | -4.82E-03 | 0.137 | 0.142 | 0.970 | 7.71E-02 | 0.139 | 0.141 | 0.910 | -4.59E-03 | 0.138 | 0.147 | 0.970 |
|     | 0.5      | 0.5        | -1.32E-03 | 0.107 | 0.107 | 0.948 | 9.98E-02 | 0.108 | 0.111 | 0.866 | 1.42E-03  | 0.108 | 0.107 | 0.940 |
|     |          | 0.5        | 4.12E-03  | 0.128 | 0.130 | 0.952 | 1.00E-01 | 0.124 | 0.127 | 0.868 | 7.45E-03  | 0.130 | 0.136 | 0.954 |
|     | 0.8      | 0.8        | 4.06E-03  | 0.086 | 0.087 | 0.950 | 6.06E-02 | 0.077 | 0.082 | 0.904 | 1.08E-02  | 0.087 | 0.087 | 0.950 |
|     |          | 0.8        | -5.57E-03 | 0.129 | 0.128 | 0.952 | 5.14E-02 | 0.120 | 0.119 | 0.916 | 1.80E-03  | 0.130 | 0.135 | 0.950 |

# Lymphatic filariasis(LF) data, background

- 림프사상충 연구 목적: 두 치료법 DEC+ALB(new)와 DEC(standard)을 비교
- 남자 47명을 대상으로 치료 후 1년 동안 주기적으로 기생충이 박멸됐는지를 관찰
  - DEC+ALB: 22명, DEC: 25명
  - 군집: 남자, 군집 안에 있는 subunit: 성충(adult worm) 집(nest)의 개수: 1~5개
  - 성충 집의 개수: 78
- 성충 집이 완전히 제거 될 때까지의 시간(days)을 측정
  - not person-specific time but nest-specific time
  - 초음파 관찰 시점: 7, 14, 30, 45, 60, 90, 180, 270, 360
  - 성충 집의 개수가 많을수록 덜 효과적
  - 실제 1개: 81.8%, 2개: 62.5%, 3개: 50%, 4 or 5: 33.3%가 박멸됨
  - 군집의 크기가 informative!
- 공변량: cluster-specific covariates
  - 치료법: DEC+ALB=0, DEC only=1
  - 나이: 16~66

# Estimated survival curves by the number of nests



# Parameter estimates

| Par   | WEE   |       |        |       |       | UWE   |       |        |       |       | WCR(B=2000) |       |        |       |       |
|-------|-------|-------|--------|-------|-------|-------|-------|--------|-------|-------|-------------|-------|--------|-------|-------|
|       | EST   | SE    | 95%CI  |       | P값    | EST   | SE    | 95%CI  |       | P값    | EST         | SE    | 95%CI  |       | P값    |
| Group | 0.591 | 0.372 | -0.138 | 1.320 | 0.112 | 0.861 | 0.445 | -0.011 | 1.732 | 0.053 | 0.594       | 0.372 | -0.136 | 1.324 | 0.111 |
| Age   | 0.005 | 0.020 | -0.035 | 0.044 | 0.815 | 0.003 | 0.024 | -0.045 | 0.050 | 0.913 | 0.005       | 0.037 | -0.068 | 0.077 | 0.900 |

- 처리 효과: WEE 와 WCR 은 insignificant, UWE는 significant. 즉, DEC 에 ALB 를 추가해도 박멸할 때까지 걸리는 시간이 다르지 않음
  - 처리효과의 계수가 양수이므로 ALB는 성충 집이 박멸될 때까지의 시간을 연장하는 효과가 있음
  - Williamson 등(2008):  $\hat{\beta}_1=0.585$  ( $p\text{값}=0.344$ )
  - Zhang & Sun(2010):  $\hat{\beta}_1=-0.267$  ( $p\text{값}=0.27$ )
- 연령 효과는 세 가지 방법 모두 insignificant!
  - Williamson 등(2008):  $\hat{\beta}_2=0.0047$  ( $p\text{값}=0.80$ )
  - Zhang & Sun(2010):  $\hat{\beta}_2=-0.088$  ( $p\text{값}<0.0001$ )

# Estimated cumulative hazard rates

|      | Estimated hazard rate |       |       | Estimated cumulative hazard rate |       |       |
|------|-----------------------|-------|-------|----------------------------------|-------|-------|
| Days | WEE                   | UWE   | WCR   | WEE                              | UWE   | WCR   |
| 7    | 0.105                 | 0.070 | 0.105 | 0.105                            | 0.070 | 0.105 |
| 14   | 0.020                 | 0.020 | 0.020 | 0.125                            | 0.090 | 0.125 |
| 30   | 0.063                 | 0.031 | 0.063 | 0.188                            | 0.121 | 0.188 |
| 45   | 0.094                 | 0.077 | 0.094 | 0.282                            | 0.198 | 0.282 |
| 60   | 0.098                 | 0.092 | 0.098 | 0.380                            | 0.290 | 0.379 |
| 90   | 0.120                 | 0.116 | 0.120 | 0.500                            | 0.405 | 0.499 |
| 180  | 0.222                 | 0.107 | 0.223 | 0.722                            | 0.512 | 0.722 |
| 270  | 0.068                 | 0.062 | 0.068 | 0.790                            | 0.574 | 0.790 |
| 360  | 0.104                 | 0.046 | 0.104 | 0.893                            | 0.620 | 0.894 |

# Concluding remarks

- 군집 구간중도절단 된 자료에서 생존시간이 군집의 크기에 의존할 때,
  - MM로부터 가중 추정 방법(WEE)
  - WCR 방법
- 모의실험 결과에 의하면 군집의 크기가
  - non-informative:  $WEE \approx UWE \approx WCR$
  - informative:  $WEE \approx WCR > UWE$
- LF data
  - nest의 개수에 따라 생존함수 추정량이 다름
  - 처리효과(DEC+ALB vs. DEC)가 유의하게 다르지는 않음
  - 연령 효과는 없음

**Thank you!**