Semiparametric Models for Multivariate Panel Count Data

KyungMann Kim

University of Wisconsin-Madison
kmkim@biostat.wisc.edu

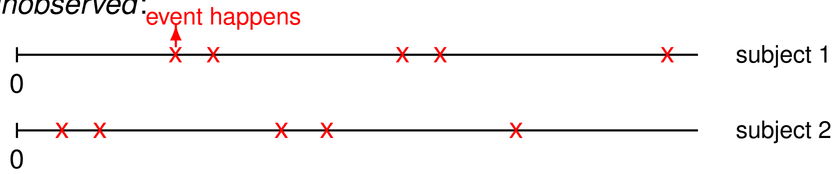
2 April 2015

Outline

Observation of the cumulative event counts for an individual at a random number of time points; both the number and the time points may differ across individuals

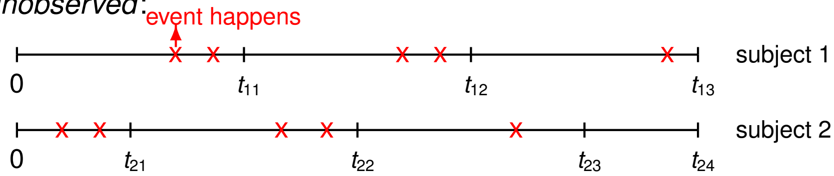
Observation of the cumulative event counts for an individual at a random number of time points; both the number and the time points may differ across individuals

Unobserved:

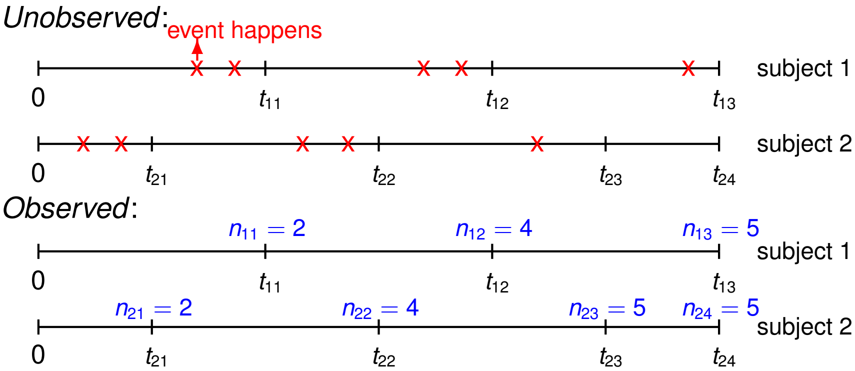


Observation of the cumulative event counts for an individual at a random number of time points; both the number and the time points may differ across individuals

Unobserved:



Observation of the cumulative event counts for an individual at a random number of time points; both the number and the time points may differ across individuals



Skin Cancer Chemoprevention Trial: Design

- Randomized, placebo-controlled, double-blind clinical trial
- Objective: Is difluoromethylornithine (DFMO) effective in preventing recurrence of non-melanoma skin cancers (NMSCs)?
- 291 subjects with prior skin cancer randomized to either DFMO or placebo
- Subjects assessed every six months for the new skin cancers until the end of study
- Primary endpoint: NMSC recurrence rate

$$\frac{\text{total number of new NMSCs}}{\text{person-year follow-up}}$$

Skin Cancer Chemoprevention Trial: Results

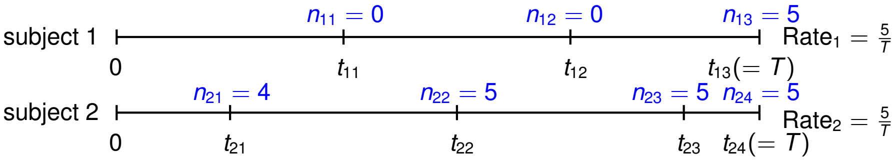
- Marginal DFMO effect on reducing the NMSC recurrence rate from 0.60 on placebo to 0.43 on DFMO based on a Poisson regression model (p-value=0.062)

Deficiencies in the Analysis

- Loss of information about longitudinal observations
- Correlation between basal and squamous cell carcinoma
- Significant DFMO effect on preventing basal cell carcinoma (BCC) (p-value=0.030)
- But not for squamous cell carcinoma (SCC) (p-value=0.565)

Deficiencies in the Analysis

- Loss of information about longitudinal observations



- Correlation between basal and squamous cell carcinoma
- Significant DFMO effect on preventing basal cell carcinoma (BCC) (p-value=0.030)
- But not for squamous cell carcinoma (SCC) (p-value=0.565)

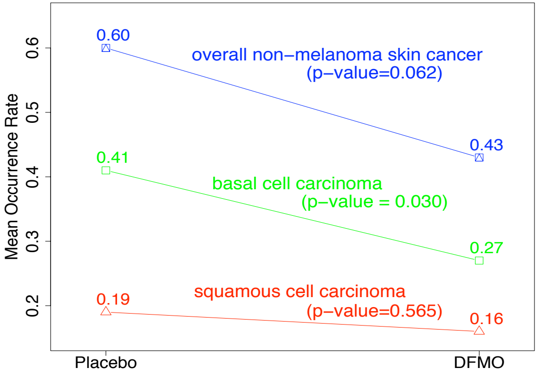
Deficiencies in the Analysis

- Loss of information about longitudinal observations
- Correlation between basal and squamous cell carcinoma
- Significant DFMO effect on preventing basal cell carcinoma (BCC) (p-value=0.030)
- But not for squamous cell carcinoma (SCC) (p-value=0.565)

Deficiencies in the Analysis

- Loss of information about longitudinal observations
- Correlation between basal and squamous cell carcinoma
- Significant DFMO effect on preventing basal cell carcinoma (BCC) (p-value=0.030)
- But not for squamous cell carcinoma (SCC) (p-value=0.565)

Poisson Analysis



Alternative Analysis

Panel count data structure

$+$

Possibly different effect of DFMO on different skin cancer types

$\Downarrow$

Data should be analyzed as the bivariate panel count data

Alternative Analysis

Panel count data structure

$$+$$

Possibly different effect of DFMO on different skin cancer types

$$\Downarrow$$

Data should be analyzed as the bivariate panel count data

Previous Work

- Univariate panel count data
- Bivariate recurrent event time data

Notations

- $N(t)$, the number of events observed between time 0 and t as a nonhomogeneous Poisson process
- Intensity function*

$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} \Pr(N(t+h) - N(t) = 1)$$

- Mean function†

$$\Lambda(t) = E[N(t)] = \int_0^t \lambda(s) ds$$

*Different from hazard function in survival analysis

†Different from cumulative hazard function in survival analysis

Univariate Panel Count Data

Inference on the intensity function $\lambda(t)$

- Parametric Poisson-Gamma frailty model (Thall, 1988)
- Nonparametric model (Thall and Lachin, 1988)
- Semiparametric Poisson-Gamma frailty model (Staniswalis *et al.*, 1997)

Univariate Panel Count Data

Inference on the mean function $\Lambda(t)$

- Nonparametric maximum pseudo-likelihood estimator (NPMPLE) (Sun and Kalbfleisch, 1995; Wellner and Zhang, 2000)

$$\hat{\Lambda}_\ell = \hat{\Lambda}(s_\ell) = \max_{r \leq \ell} \min_{u \geq \ell} \frac{\sum_{v=r}^{u} w_v \bar{n}_v}{\sum_{v=r}^{u} w_v}$$

  where $w_v = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \mathrm{I}\{t_{ij} = s_v\}$,
  $\bar{n}_v = \frac{1}{w_v} \sum_{i=1}^{n} \sum_{j=1}^{k_i} n_{ij} \mathrm{I}\{t_{ij} = s_v\}$, $n_{ij} = N(t_{ij})$ is the cumulative event counts up to time t_{ij} of subject i, k_i is the number of follow-up visits, and $\{s_v\}_{v=1}^{L}$ is the unique, ordered $\{t_{ij}; j = 1, ..., k_i, i = 1, ..., n\}$
- Semiparametric model (Zhang, 2002)
  Assuming mean function $\Lambda(t) = \Lambda_0(t) \exp(\beta' Z(t))$
- Gamma frailty model (Zhang and Jamshidian, 2003)
  Assuming the conditional mean function $\Lambda(t|\gamma) = \gamma\Lambda_0(t)$

Univariate Panel Count Data

Inference on the mean function $\Lambda(t)$

- Nonparametric maximum pseudo-likelihood estimator (NPMPLE) (Sun and Kalbfleisch, 1995; Wellner and Zhang, 2000)
- Semiparametric model (Zhang, 2002)
  Assuming mean function $\Lambda(t) = \Lambda_0(t)\exp(\beta' Z(t))$
- Gamma frailty model (Zhang and Jamshidian, 2003)
  Assuming the conditional mean function $\Lambda(t|\gamma) = \gamma\Lambda_0(t)$
- Nonparametric test for univariate panel count data based on NPMPLE (Sun and Fang, 2003)

Bivariate Recurrent Event Time Data

- Parametric model for bivariate recurrent event time data (Abu-Libdeh *et al.* 1990)
  - $N_1(t)$ and $N_2(t)$ are two counting processes
  - Two types of random effects:
    Subject level random effect $\gamma \sim Gamma(\alpha, \nu)$
    Process random effects $\boldsymbol{\xi} = (\xi_1, \xi_2) \sim$ Dirichlet $(\upsilon = (\upsilon_1, \upsilon_2))$
  - Given $(\gamma, \boldsymbol{\xi})$, $N_1(t)$ and $N_2(t)$ are independent nonhomogeneous Poisson processes

$$\lambda_p(t|\gamma, \xi_p) = \gamma \xi_p \delta t^{\delta - 1} \exp(\beta' Z), \;\; p = 1, 2$$

  - Obtain maximum likelihood estimators of $(\beta, \alpha, \nu, \upsilon_1, \upsilon_2)$
- Not applicable to the panel count data since event times are unknown

Semiparametric Frailty Models

- Frailty models
- Estimation procedures
- Statistical inference on regression parameters

General Model Assumptions

- Frailty variable $\gamma \sim g(\eta)$
  - Subject heterogeneity
  - Dependence between processes comes from the common frailty variable
- Given γ, $N_1(t)$ and $N_2(t)$ are independent nonhomogeneous Poisson processes with conditional mean functions

$$\Lambda_1(t|\gamma) = \gamma\Lambda_{10}(t)\exp(\beta_{\mathbf{1}}' Z) \text{ and } \Lambda_2(t|\gamma) = \gamma\Lambda_{20}(t)\exp(\beta_{\mathbf{2}}' Z)$$

Likelihood Function

- The complete log-likelihood function of $(\Lambda_{10}, \Lambda_{20}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \eta)$ is simplified as

$$\ell_n(\Lambda_{10}, \Lambda_{20}, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \left\{ \sum_{p=1}^{2} [n_{ijp} \log \Lambda_{p0ij} + n_{ijp} \boldsymbol{\beta}_{\boldsymbol{p}}' Z - \gamma_i \Lambda_{p0ij} e^{\boldsymbol{\beta}_{\boldsymbol{p}}' Z}] \right\} + \sum_{i=1}^{n} h_i(\eta)$$

  - $n_{ijp} = N_p(t_{ij})$, the cumulative type p event counts up to time t_{ij}
  - $\Lambda_{p0ij} = \Lambda_{p0}(t_{ij})$, the baseline mean function of type p event at t_{ij}

Estimating Procedures

EM algorithm to obtain the MLEs

- E-step: same as in the nonparametric models
- M-step: involves the profile-likelihood type estimators of $(\Lambda_{10}(t), \Lambda_{20}(t), \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$ and MLE of η
  step 1: Given $(\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)})$, obtain $(\hat{\Lambda}_{10}^{(1)}, \hat{\Lambda}_{20}^{(1)})$
  step 2: Given $(\hat{\Lambda}_{10}^{(k)}, \hat{\Lambda}_{20}^{(k)})$, obtain $(\boldsymbol{\beta}_1^{(k)}, \boldsymbol{\beta}_2^{(k)})$
  step 3: Repeat steps 1 and 2 until convergence

Models Investigated

- Bivariate Poisson-Gamma frailty models
- Bivariate Poisson-lognormal frailty models

Bivariate Poisson-Gamma Frailty Models

EM algorithm: E-step

- Posterior expectation of γ_i given current estimates $(\tilde{\Lambda}_{10}, \tilde{\Lambda}_{20}, \tilde{\boldsymbol{\beta}}_{\mathbf{1}}, \tilde{\boldsymbol{\beta}}_{\mathbf{2}}, \tilde{\alpha})$ given by

$$\tilde{\gamma}_i = \frac{\sum_{j=1}^{k_i}(n_{ij1} + n_{ij2}) + \tilde{\alpha}}{\sum_{j=1}^{k_i}(\tilde{\Lambda}_{10ij} e^{\tilde{\boldsymbol{\beta}}_{\mathbf{1}}{}' Z} + \tilde{\Lambda}_{20ij} e^{\tilde{\boldsymbol{\beta}}_{\mathbf{2}}{}' Z}) + \tilde{\alpha}}$$

- Posterior expectation of $\log \gamma_i$, $\widetilde{\log \gamma_i}$, is calculated by Monte-Carlo method

Bivariate Poisson-Gamma Frailty Models

EM algorithm: M-step (1)

Profile likelihood type estimators of $(\Lambda_{10}, \Lambda_{20}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$

- Step 1: Given $(\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)})$,

$$\hat{\Lambda}_{10}^{(1)}(s_\ell) = \hat{\Lambda}_{10\ell}^{(1)} = \max_{r \le \ell} \min_{u \ge \ell} \frac{\sum_{r \le v \le u} \omega_{1v} \bar{n}_{v1}}{\sum_{r \le v \le u} \omega_{1v}}$$

$$\hat{\Lambda}_{20}^{(1)}(s_\ell) = \hat{\Lambda}_{20\ell}^{(1)} = \max_{r \le \ell} \min_{u \ge \ell} \frac{\sum_{r \le v \le u} \omega_{2v} \bar{n}_{v2}}{\sum_{r \le v \le u} \omega_{2v}}$$

  - $\bar{n}_{v1} = \frac{1}{\omega_{1v}} \sum_{i=1}^{n} \sum_{j=1}^{k_i} n_{ij1} \mathrm{I}\{t_{ij} = s_v\}$ $\bar{n}_{v2} = \frac{1}{\omega_{2v}} \sum_{i=1}^{n} \sum_{j=1}^{k_i} n_{ij2} \mathrm{I}\{t_{ij} = s_v\}$
  - $\omega_{1v} = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \tilde{\gamma}_i e^{\boldsymbol{\beta}_1^{(0)\prime} Z} \mathrm{I}\{t_{ij} = s_v\}$

    $\omega_{2v} = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \tilde{\gamma}_i e^{\boldsymbol{\beta}_2^{(0)\prime} Z} \mathrm{I}\{t_{ij} = s_v\}$

Bivariate Poisson-Gamma Frailty Models

EM algorithm: M-step (2)

- step 2: Given $\hat{\Lambda}_{10\ell}^{(k)}, \hat{\Lambda}_{20\ell}^{(k)}$, obtain MLE of $(\boldsymbol{\beta}_1^{(k)}, \boldsymbol{\beta}_2^{(k)})$ by maximizing

$$\sum_{i=1}^{n}\sum_{j=1}^{k_i}[n_{ij1}\log\Lambda_{10ij}^{(1)} + n_{ij1}\boldsymbol{\beta}_1' Z - \tilde{\gamma}_i\Lambda_{10ij}^{(1)} e^{\boldsymbol{\beta}_1' Z}]$$

$$\sum_{i=1}^{n}\sum_{j=1}^{k_i}[n_{ij2}\log\Lambda_{20ij}^{(1)} + n_{ij2}\boldsymbol{\beta}_2' Z - \tilde{\gamma}_i\Lambda_{20ij}^{(1)} e^{\boldsymbol{\beta}_2' Z}]$$

- step 3: Solve $(\hat{\Lambda}_{10\ell}^{(k)}, \hat{\Lambda}_{20\ell}^{(k)}, \hat{\boldsymbol{\beta}}_1^{(k)}, \hat{\boldsymbol{\beta}}_2^{(k)})$ iteratively until convergence

MLE of α:

$$\hat{\alpha} = \arg\max_{\alpha>0}[\alpha\sum_{i=1}^{n}(\widetilde{\log\gamma_i} - \tilde{\gamma}_i) + n\alpha\log\alpha - n\log\Gamma(\alpha)]$$

Bivariate Poisson-Lognormal Frailty Models

EM algorithm: E-step

- Posterior expectations of $(\gamma_{1i}, \gamma_{2i})$, denoted by $(\tilde{\gamma}_{1i}, \tilde{\gamma}_{2i})$, and $(\log\gamma_{1i}, \log\gamma_{2i})$, denoted by $(\widetilde{\log\gamma_{1i}}, \widetilde{\log\gamma_{2i}})$, are calculated by importance sampling method

Bivariate Poisson-Lognormal Frailty Models

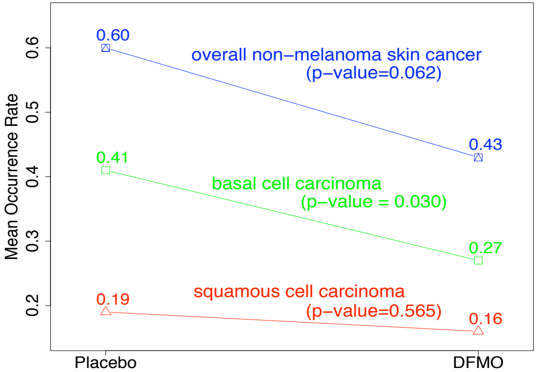
EM algorithm: M-step

- Profile likelihood type estimates of $(\Lambda_{10}, \Lambda_{20}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$
- Similar procedures as in the Poisson-Gamma model
- MLE of $(\sigma_1^2, \sigma_2^2, \rho)$
  Same estimators as mentioned in the nonparametric Poisson-lognormal frailty model

Statistical Inference

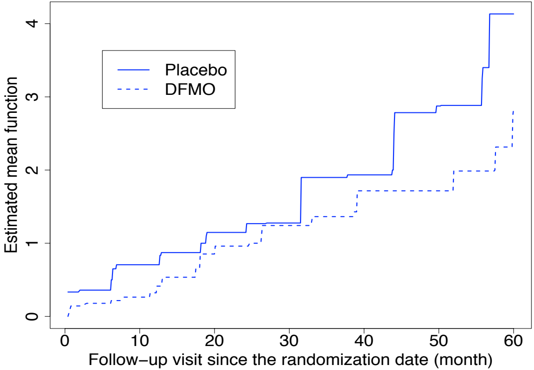
- Bootstrap is used to obtain the estimated variance of the estimators
- Wald test for the regression parameters based on bootstrap estimated variance

Recall: Poisson Analysis



Univariate Models

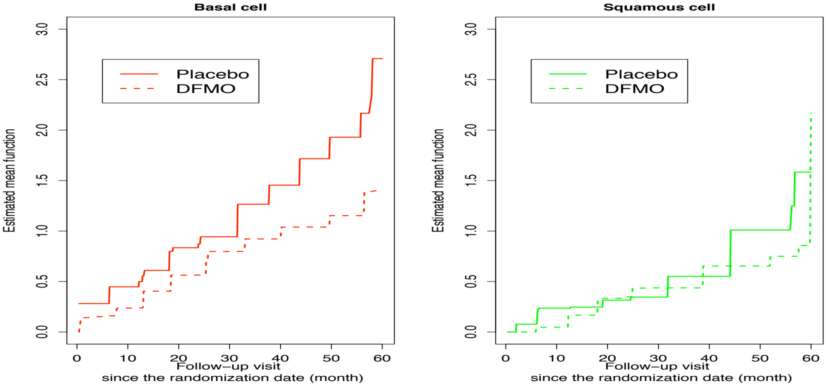
There was marginal treatment effect of DFMO on prevention the recurrence of NMSC (*p*-value=0.098*)



*Based on the test for univariate panel count data (Sun and Fang, 2003)

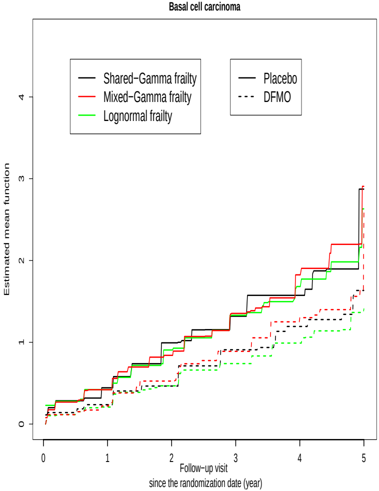
Two Separate Univariate Models

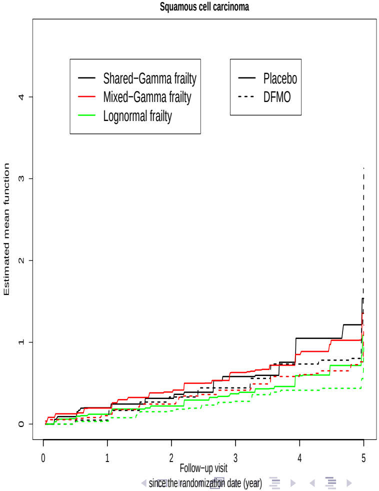
There was significant treatment effect of DFMO on preventing the recurrence of BCC (p-value=0.034*), but not for SCC (p-value=0.585*)



*Based on the test for univariate panel count data (Sun and Fang, 2003)

Mean Functions for BCC and SCC





Nonparametric Tests

	p-values			
Models	NMSC	Global	BCC	SCC
Univariate	0.098^{*}		0.033^{*}	0.585^{*}
Gamma		$0.130^{\dagger}$	0.044^{*}	0.626^{*}
Lognormal		$0.114^{\dagger}$	0.190^{*}	0.233^{*}
Mixed Gamma	0.130^{*}			

* *p*-values are calculated based on the univariate test (Sun and Fang, 2003)

† *p*-values are calculated based on the bivariate test

Bivariate Poisson-Gamma Frailty Models

Mean Functions

- Let $N_1(t)$ and $N_2(t)$ represent the number of BCCs and SCCs, respectively, with marginal mean functions

$$E[N_1(t)|Z] = \Lambda_{10}(t)\exp(\beta_{11}Z_1 + \beta_{12}Z_2)$$

$$E[N_2(t)|Z] = \Lambda_{20}(t)\exp(\beta_{21}Z_1 + \beta_{22}Z_2)$$

- Z_1, the treatment group (placebo/DFMO)
- Z_2, the logarithm of the previous tumor rate

Bivaraite Poisson-Gamma Frailty Models

Statistical Inference

Marginal treatment effect on decreasing recurrence of BCC and SCC after adjusting for the baseline tumor rates

Tumor type	Covariate	Estimate	Bootstrap SE	p-value
BCC	Z_1	-0.282	0.149	0.058
	Z_2	0.138	0.094	0.142
SCC	Z_1	-0.033	0.336	$>$0.5
	Z_2	0.218	0.160	0.173

Summary

Semiparametric models of the mean functions for a bivariate counting process are proposed

- Methods of analyzing bivariate panel count data
- No specific form of mean functions assumed
- Providing estimation of two mean functions simultaneously
- Correlation between processes can be derived
- Event-type specific covariate effects assumed in the semiparametric models

Poisson-Gamma *vs.* Poisson-Lognormal Frailty

Gamma frailty	Lognormal frailty
One-variable model	Two-variable model
	Frailty variables can be negatively correlated
Easier to implement	More complicate
Shorter computing time	Longer computing time
	Easier to genearlized when there are mroe than two event types of interest