

PENDEKATAN KEKAR UNTUK MODEL BERSAMA (*JOINT MODEL*) ATAS DASAR SEBARAN t (*A Robust Approach for Joint Model Based on t Distribution*)

Indahwati¹, Aunuddin¹, Khairil Anwar Notodiputro¹, I Gusti Putu Purnaba²

¹Departemen Statistika, FMIPA IPB

²Departemen Matematika, FMIPA IPB

email : indahwati_43@yahoo.co.id

Existing methods for joint modeling are usually based on normality assumption of random effects and intra subject errors. We propose a joint model based on t distribution of the intra subject errors to improve robustness of the estimation. Our model consists of two submodels: a mixed linear mixed effects model for the longitudinal data, and a generalized linear model for continuous/binary primary response. The proposed method is evaluated by means of simulation studies as well as application to HIV data.

Keywords: joint modeling, longitudinal data, robust, t distribution

PENDAHULUAN

Salah satu pendekatan yang dilakukan untuk memodelkan hubungan antara peubah penjelas longitudinal dengan peubah respon skalar - yang dalam kepustakaan seringkali disebut *primary endpoint* - adalah pemodelan bersama (*joint modeling*). Pendekatan ini seringkali digunakan untuk memodelkan hubungan antara data longitudinal (sebagai peubah penjelas) dengan data daya tahan hidup (sebagai peubah respon), diantaranya dapat dijumpai dalam Henderson *et al.* (2000), serta Tsiatis dan Davidian (2004). Pemodelan bersama dengan respon primer berupa data biner diantaranya dapat dijumpai dalam Zhang dan Lin (1999), Li *et al.* (2004, 2007a, 2007b), serta Horrocks dan Heuvel (2009).

Pendekatan paling populer untuk model bersama (*joint model*) adalah model berbagi parameter (*shared parameter model*). Prinsip umum dari pendekatan ini adalah penggabungan dua model, yaitu submodel-1 yang diasumsikan mengikuti model linier campuran (*linear mixed model*) - mungkin setelah ditransformasi - untuk memodelkan data longitudinal, serta submodel-2 yaitu model regresi primer yang diasumsikan mengikuti model linier terampat (*generalized linear model*) untuk respon primer yang mengikuti sebaran keluarga eksponensial, atau model proporsional hazard untuk respon primer yang berupa data daya tahan hidup. Dalam pendekatan model bersama, peubah respon dalam model regresi primer bergantung pada kovariat longitudinal melalui pengaruh acak spesifik subyek. Dengan kata lain, pengaruh acak yang

dihasilkan dari model campuran (submodel-1) menjadi peubah bebas pada model regresi primer (submodel-2). Namun karena pengaruh acak tak teramati, maka pendekatan "naive" dengan cara mensubstitusi langsung nilai dugaan OLS (*Ordinary Least Squares*) setiap subyek dari submodel-1 ke dalam submodel-2 sebagai peubah penjelas akan menghasilkan nilai dugaan parameter model regresi primer yang berbias, khususnya yang mengukur pengaruh kovariat longitudinal terhadap peubah respon primer (Zhang dan Lin 1999; Wang *et al.* 2000). Metode pendugaan parameter dalam pemodelan bersama umumnya menggunakan metode kemungkinan maksimum atau metode Bayes.

Selama ini metode pendugaan parameter pada model bersama didasarkan atas asumsi bahwa hasil pengukuran longitudinal mengikuti model linier campuran dengan pengaruh acak dan galat intra-subyek menyebar normal. Namun adakalanya asumsi kenormalan tersebut tidak terpenuhi. Dalam penelitian ini akan diterapkan pendekatan lain yang lebih kekar (*robust*) terhadap sebaran galat intra subyek yang tidak menyebar normal.

Selama tiga dekade terakhir, sebaran peubah ganda t (*multivariate t distribution*) telah dikenal sebagai salah satu generalisasi yang berguna sebagai pendekatan kekar, terutama terhadap pencilan, misalnya dalam model regresi linier (Zellner 1976; Lange *et al.* 1989) dan model linier campuran (Pinheiro *et al.* 2001; Song *et al.* 2007; Wang & Fan 2010). Kotz dan Nadarajah (2004), serta Nadarajah dan Kotz (2008) menyatakan bahwa dibandingkan sebaran normal ganda,

sebaran peubah ganda t menawarkan alternatif yang lebih fleksibel berkaitan dengan sebaran data di dunia riil, karena mempunyai ekor yang lebih realistis. Dalam konteks pemodelan bersama dengan respon primer berupa data daya tahan hidup, pendekatan kekar terhadap pencilan dikaji oleh Li *et al.* (2009) dengan mengasumsikan galat intra subyek menyebar t -student.

Penelitian ini bertujuan untuk (1) membuat formulasi model bersama dengan galat intra subyek menyebar t , (2) mengkaji penggunaan sebaran t untuk galat intra subyek sebagai pendekatan yang lebih kekar terhadap asumsi kenormalan melalui kajian simulasi. Kajian empiris akan diterapkan terhadap data kematian penderita HIV/AIDS yang dihubungkan dengan hasil pengukuran longitudinal jumlah sel CD4⁺ pasien beserta kovariat lainnya.

MODEL BERSAMA DENGAN GALAT INTRA SUBYEK MENYEBAR t

Misalkan Y_i adalah peubah respon dari subyek ke- i , $i = 1, 2, \dots, n$ pada model regresi primer, \mathbf{Z}_i adalah vektor kovariat berdimensi $p \times 1$, dan $\mathbf{W}_i = (\mathbf{W}_{i1}, \dots, \mathbf{W}_{im_i})'$ merupakan kovariat longitudinal yang diukur pada m_i titik waktu $(t_{i1}, \dots, t_{im_i})$. Yang menjadi perhatian adalah memodelkan pengaruh kovariat \mathbf{Z}_i dan trajektori dari kovariat longitudinal \mathbf{W}_i terhadap peubah respon Y_i . Data longitudinal pada submodel-1 dimodelkan sebagai berikut:

$$\mathbf{W}_{ij} = \mathbf{X}_{ij}'\mathbf{b}_i + \varepsilon_{ij}$$

$$\mathbf{b}_i \sim N_p(\boldsymbol{\eta}, \boldsymbol{\Sigma}) \text{ dan } \varepsilon_{ij} \sim t(0, \sigma_\varepsilon^2, \nu)$$

\mathbf{X}_{ij} adalah vektor kovariat berdimensi $q \times 1$, \mathbf{b}_i adalah vektor pengaruh acak, serta ε_{ij} adalah galat intra subyek. Adapun ν adalah parameter derajat bebas dari sebaran t -student yang bisa diduga, atau ditetapkan untuk penyederhanaan dan mengurangi beban komputasi. Jika $\mathbf{X}_{ij} = (1, t_{ij})'$, komponen b_{0i} dan b_{1i} dari \mathbf{b}_i dapat diinterpretasikan sebagai nilai baseline dan laju perubahan X terhadap Y bagi subyek ke- i (Zhang dan Lin 1999).

Sebaran bersyarat $Y_i | \mathbf{b}_i$ diasumsikan berasal dari sebaran keluarga eksponen (McCullagh dan Nelder 1989) dengan nilai tengah $E(Y_i | \mathbf{b}_i) = \mu_i$ dan ragam $Var(Y_i | \mathbf{b}_i) = \phi a_i^{-1} \nu(\mu_i)$, dimana ϕ adalah parameter dispersi, a_i merupakan pembobot prior, dan $\nu(\cdot)$ adalah fungsi ragam. Pengaruh dari $(\mathbf{W}_i, \mathbf{Z}_i)$ terhadap Y_i dimodelkan melalui $(\mathbf{b}_i, \mathbf{Z}_i)$ menggunakan model linier terampat:

$$g(\mu_i) = \beta_0 + \mathbf{Z}_i' \boldsymbol{\beta}_1 + \mathbf{b}_i' \boldsymbol{\beta}_2$$

sedangkan $g(\cdot)$ adalah fungsi hubung. Lebih lanjut diasumsikan bahwa Y_i dan \mathbf{W}_i bebas bersyarat

terhadap \mathbf{b}_i . Dalam hal ini inferensi terutama ditujukan terhadap parameter regresi $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$.

Berdasarkan Model Berbagi Parameter (*Shared Parameter Model*), hubungan antara \mathbf{W}_i dan Y_i dimodelkan melalui fungsi kepekatan bersama $f(\mathbf{w}_i, y_i)$. Dengan asumsi $Y_i | \mathbf{b}_i$ dan $\mathbf{W}_i | \mathbf{b}_i$ saling bebas, serta $\mathbf{W}_{ij} | \mathbf{b}_i, j=1, 2, \dots, m_i$ saling bebas, maka:

$$\begin{aligned} f(\mathbf{w}_i, y_i) &= \int f(\mathbf{w}_i, y_i, \mathbf{b}_i) d\mathbf{b}_i \\ &= \int f(\mathbf{w}_i, y_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\ &= \int f(\mathbf{w}_i | \mathbf{b}_i) f(y_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\ &= \int \prod_{j=1}^{m_i} f(w_{ij} | \mathbf{b}_i) f(y_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \end{aligned}$$

Misalkan $\boldsymbol{\gamma} = (\boldsymbol{\eta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$. Fungsi kemungkinan dan log-kemungkinan bagi $\Theta = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ adalah:

$$\begin{aligned} L(\Theta | \mathbf{w}_i, y_i) &= \prod_{i=1}^n \int \prod_{j=1}^{m_i} f(w_{ij} | \mathbf{b}_i) f(y_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\ l(\Theta | \mathbf{w}_i, y_i) &= \sum_{i=1}^n \log \int \prod_{j=1}^{m_i} f(w_{ij} | \mathbf{b}_i) f(y_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \end{aligned}$$

Untuk peubah kontinu yang menyebar normal, maka $g(\cdot)$ merupakan fungsi hubung identitas, sehingga $\mu_i = \beta_0 + \mathbf{Z}_i' \boldsymbol{\beta}_1 + \mathbf{b}_i' \boldsymbol{\beta}_2$ atau $Y_i = \beta_0 + \mathbf{Z}_i' \boldsymbol{\beta}_1 + \mathbf{b}_i' \boldsymbol{\beta}_2 + \varepsilon_i$. Jika $\varepsilon_i \sim N(0, \sigma^2)$, maka bersyarat pada \mathbf{b}_i , Y_i akan menyebar normal dengan nilai tengah $E(Y_i | \mathbf{b}_i) = \beta_0 + \mathbf{Z}_i' \boldsymbol{\beta}_1 + \mathbf{b}_i' \boldsymbol{\beta}_2$ dan ragam $Var(Y_i | \mathbf{b}_i) = \sigma^2$. Dengan demikian secara lengkap fungsi log-kemungkinan bagi $\Theta = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ dapat dituliskan sebagai:

$$\begin{aligned} l(\Theta | \mathbf{W}_i, Y_i) &= \sum_{i=1}^n \log \left\{ \prod_{j=1}^{m_i} \frac{\Gamma\{(v+1)/2\}}{\sqrt{\pi v \sigma_\varepsilon^2} \Gamma(\frac{1}{2})} \left[1 + \frac{1}{\sigma_\varepsilon^2 v} (w_{ij} - \mathbf{X}_{ij}' \boldsymbol{\eta})^2 \right]^{-(v+1)/2} \right. \\ &\quad \times \left. \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \beta_0 + \mathbf{Z}_i' \boldsymbol{\beta}_1 + \mathbf{b}_i' \boldsymbol{\beta}_2)^2 \right] \right\} \\ &\quad \times \left[\frac{1}{\sqrt{2\pi} |\boldsymbol{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{b}_i - \boldsymbol{\eta})' \boldsymbol{\Sigma}^{-1} (\mathbf{b}_i - \boldsymbol{\eta}) \right] \right] d\mathbf{b}_i \end{aligned}$$

Dengan mengambil nilai parameter ν tetap, penduga parameter *joint model* dapat diperoleh dengan memaksimalkan fungsi log-kemungkinan di atas terhadap parameter $\Theta = (\boldsymbol{\gamma}, \boldsymbol{\beta})$, sedangkan $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ dan $\boldsymbol{\gamma} = (\boldsymbol{\eta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$.

KAJIAN SIMULASI

Untuk melihat pengaruh pelanggaran asumsi sebaran normal dari galat intra subyek terhadap sifat-sifat penduga parameter model bersama, dilakukan kajian simulasi dengan kombinasi beberapa kondisi. Secara garis besar berikut adalah tahapan simulasi yang diajukan:

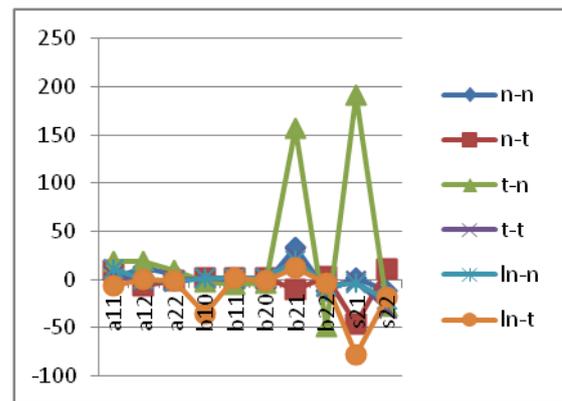
1. Pembangkitan Data
 - a. Ukuran contoh (subyek) ditetapkan sebesar $n=100$.
 - b. Untuk data longitudinal pada submodel-1, data dibangkitkan mengikuti model koefisien acak $W_{ij} = X_{ij}'b_i + \varepsilon_{ij}$. Matriks $X_{ij} = (1, t_{ij})'$, $j = 1, \dots, m_i$, adalah *fixed* dengan $m_i = 4$ untuk setiap i , diamati pada $t = 1, 3, 6, 9$. Pengaruh acak $b_i = (b_{0i}, b_{1i})$ dibangkitkan dari sebaran bivariat normal dengan $E(b_{0i}, b_{1i}) = (1, 1)$, $\text{var}(b_{1i}) = \sigma_{b1}^2 = \sigma_{b2}^2 = \text{var}(b_{2i}) = 2$ dan $\text{cor}(b_{1i}, b_{2i}) = r_{12} = 0.7$. Adapun galat intra subyek ε_{ij} dibangkitkan dari tiga sebaran yaitu: (i) sebaran normal dengan nilai tengah 0 dan ragam $\sigma_\varepsilon^2 = 2$, (ii) sebaran- t dengan derajat bebas 3 untuk mewakili sebaran yang berekor panjang (iii) sebaran lognormal untuk mewakili sebaran yang menjulur.
 - c. Untuk submodel-2, data dibangkitkan mengikuti model $Y_i = b_i'\beta + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$. Kovariat lain pada model di atas ditiadakan mengingat tujuan utama pemodelan adalah untuk mengetahui apakah ada hubungan antara profil longitudinal dengan peubah respon primernya, yang dalam hal ini diukur oleh parameter β . Nilai parameter β ditetapkan sebesar (2,5) dan $\sigma^2 = 2$.
2. Pemodelan dengan Model Bersama dan Pendugaan Parameter
Data bangkitan dari kedua submodel pada butir (1) selanjutnya dimodelkan dengan model bersama $f(W_{ij}, Y_i)$. Untuk melihat pengaruh ketaknormalan galat intra subyek, maka ketiga sebaran galat intra subyek butir (1) akan dibandingkan bagaimana jika diasumsikan galat intra subyeknya menyebar normal dan t . Setiap kondisi simulasi dilakukan sebanyak $R = 200$ kali.
3. Evaluasi dilakukan melalui kajian sifat-sifat dari nilai dugaan parameter meliputi rataan bias relatif dan MSE sebagai berikut:

$$ARB = \frac{1}{R} \sum_{r=1}^R \left(\frac{\hat{\theta}_r - \theta}{\theta} \right)$$

$$MSE = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta)^2$$

Terdapat sepuluh parameter yang diduga dalam pemodelan bersama ini. Efek tetap pada submodel-1 dilambangkan sebagai b_{10} dan b_{11} , sedangkan pada submodel-2 dilambangkan sebagai b_{20} , b_{21} , dan b_{22} . Adapun a_{11} , a_{12} , dan a_{12} masing-masing merupakan ragam intersep, peragam (intersep, slope), serta ragam slope dari submodel-1. Berikutnya, s_{21} merupakan ragam galat intra subyek pada submodel-1, sedangkan s_{22} adalah ragam galat submodel-2.

Ringkasan hasil simulasi untuk rataan bias relatif (ARB) disajikan pada Gambar 1 dan Tabel 1. Dari Gambar 1 dapat dilihat bahwa bias ke atas yang sangat besar terjadi untuk parameter b_{21} dan s_{21} dari galat intra subyek yang berasal dari sebaran setangkup berekor panjang (t dengan db 3) yang diasumsikan normal. Demikian pula untuk parameter b_{22} bias ke bawahnya juga cukup besar dibandingkan sebaran lainnya.



Gambar 1. Grafik ARB (%) penduga parameter model bersama dengan galat intra subyek menyebar normal, t , lognormal diasumsikan menyebar normal dan t

Pendekatan dengan sebaran t memberikan hasil yang lebih baik dengan lebih kecilnya nilai ARB untuk parameter b_{21} menjadi 23.41% dari sebelumnya 157.07% jika menggunakan pendekatan normal. Demikian pula untuk parameter s_{21} , terjadi penurunan bias dari semula 191.84% untuk pendekatan normal menjadi -1.18% jika menggunakan pendekatan sebaran t . Hal yang sama juga terjadi pada parameter b_{22} , dari nilai ARB semula -48.62% menjadi -7.51% (Tabel 1).

Untuk sebaran yang menjulur (lognormal), pendekatan sebaran t memberikan bias yang lebih kecil untuk sebagian besar parameter dibandingkan pendekatan sebaran normal. Namun untuk parameter b_{10} dan s_{21} terjadi bias ke bawah yang

cukup besar jika menggunakan pendekatan sebaran t (Gambar 1 dan Tabel 1).

Dari Tabel 1 dapat dilihat bahwa penduga parameter b_{21} berbias agak besar untuk semua sebaran, dan biasanya sangat besar jika sebaran asalnya t diasumsikan normal, demikian pula untuk penduga parameter b_{22} . Untuk parameter s_{22} , tetap menghasilkan nilai dugaan yang berbias kecuali jika sebaran asalnya normal dan dimodelkan normal.

Nilaitengah jumlah kuadrat (MSE) hasil simulasi disajikan pada Tabel 2. Kecuali untuk galat intra subyek dari sebaran t yang diasumsikan normal, secara umum nilai MSE untuk semua kondisi simulasi bernilai kecil, terutama untuk parameter efek tetap. Dapat disimpulkan bahwa pendekatan sebaran normal untuk sebaran simetrik yang berekor panjang memberikan nilai dugaan yang tidak stabil, terutama untuk parameter b_{21} , b_{22} , dan s_{21} .

Tabel 1. Nilai ARB (%) penduga parameter model bersama dengan galat intra subyek menyebar normal, t , lognormal diasumsikan menyebar normal dan t

Sebaran	normal			t		
	normal	t	lognormal	normal	t	lognormal
Asumsi	normal			t		
a11	5.48	19.13	10.7	9.73	3.91	-6.58
a12	12.82	19.38	-0.3	-5.79	-1.75	0.44
a22	6.13	9.01	-2.67	-1.22	-0.52	-1.52
b10	-4.84	-2.78	1.96	1.09	1.57	-35.88
b11	-6.49	-4.51	2.17	1.08	0.48	1.04
b20	-6.22	-4.24	1.87	1.1	0.28	-1.44
b21	25.08	157.07	20.78	-10.11	23.41	11.96
b22	-8.7	-48.62	-6.75	3.26	-7.51	-3.98
s21	1.76	191.84	-3.27	-46.05	-1.18	-77.96
s22	-0.56	-27.09	-21.65	10.23	-30.37	-17.79

Tabel 2. Nilai MSE penduga parameter model bersama dengan galat intra subyek menyebar normal, t , lognormal diasumsikan menyebar normal dan t

Sebaran	normal			t		
	normal	t	lognormal	normal	t	lognormal
Asumsi	normal			t		
a11	0.29	2.74	0.65	0.32	0.53	0.16
a12	0.27	0.76	0.08	0.09	0.11	0.07
a22	0.18	0.37	0.09	0.08	0.09	0.09
b10	0.06	0.1	0.03	0.04	0.04	0.15
b11	0.05	0.06	0.02	0.02	0.02	0.02
b20	2.48	2.6	0.87	0.82	0.82	0.84
b21	1.14	246.08	1.74	0.2	1.52	0.21
b22	0.78	122.61	1.12	0.15	1.05	0.14
s21	0.05	19.96	0.05	0.86	0.05	2.45
s22	1.85	2.88	1.35	0.43	1.39	0.38

PENERAPAN PADA DATA KASUS HIV/AIDS

Pemodelan bersama diterapkan terhadap hasil suatu percobaan klinis untuk membandingkan kemanjuran dan keamanan dua jenis obat antiretroviral dalam menangani pasien-pasien yang gagal atau tidak toleran terhadap terapi zidovudine (AZT). Percobaan melibatkan $n = 467$ pasien terinfeksi HIV yang terdiagnosa sebagai penderita AIDS atau memiliki jumlah sel $CD4^+ \leq 300$ per ml^3 darah. Pasien dibagi secara acak untuk menerima salah satu dari dua jenis obat, yaitu didanosine (ddI) atau zalcitabine (ddC).

Banyaknya sel $CD4^+$ dicatat pada saat terlibat dalam studi ($t = 0$), dan kunjungan pada bulan ke 2, 6, 12 dan 18, sehingga maks $m_i = 5$. Data ini juga digunakan oleh Guo dan Carlin (2004) untuk pemodelan bersama data longitudinal dan data daya tahan hidup (waktu sampai terjadinya kematian) dari penderita HIV. Karena data longitudinal sangat menjulur ke kanan, dilakukan transformasi data sebelum analisis berikutnya. Transformasi akar dipilih untuk mengurangi kemenjuluran pola sebaran sekaligus untuk $CD4^+$ menstabilkan ragam, juga karena datanya merupakan data cacahan.

Peubah respon biner pada submodel-2 adalah terjadinya kematian pasien pada saat penelitian (Death = 1, NoDeath = 0), sedangkan peubah penjelasnya adalah Drug (ddI = 1, ddC = 0), Gender (male = 1, female = -1), PrevOI (AIDS diagnosis at study entry = 1, no AIDS diagnosis = -1), dan Stratum (AZT failure = 1, AZT intolerance = -1). Untuk submodel-1, sebagai peubah responnya adalah banyaknya sel CD4⁺ pada t = 0, 2, 6, 12, dan 18 bulan, sedangkan peubah penjelasnya sama dengan di atas ditambah pengaruh waktu.

Data longitudinal hasil transformasi akar banyaknya sel CD4⁺ dalam submodel-1 selanjutnya dimodelkan sebagai model linier campuran dengan persamaan sebagai berikut:

$$w_{ij} = \beta_{01} + \beta_{11}Time_{ij} + \beta_{21}Time_{ij} \times Drug_i + \beta_{31}Gender_i + \beta_{41}PrevOI_i + \beta_{51}Stratum_i + b_{0i} + b_{1i}Time_{ij} + \varepsilon_{ij},$$

$$i = 1, 2, \dots, 467 \quad j = 1, 2, \dots, m_i$$

sedangkan:

$$\mathbf{b}' = (b_{0i}, b_{1i})' \sim N_2(\boldsymbol{\eta}, \boldsymbol{\Sigma}) \quad \text{dan} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Pada model di atas, galat intra subyek tidak dimodelkan mengikuti sebaran t , karena ukuran contoh yang cukup besar, dan hasil diagnostik terhadap sisaan menunjukkan pola yang menyebar normal.

Selanjutnya peubah respon primer terjadi atau tidaknya kematian dalam masa studi untuk pasien ke- i dalam submodel-2 dimodelkan sebagai model linier terampat untuk respon biner (regresi logistik) dengan persamaan berikut:

$$P(y_i = 1 | \mathbf{b}) = \frac{\exp(A)}{1 + \exp(A)}$$

$$\text{sedangkan} \quad A = \beta_{02} + \beta_{12}Drug_i + \beta_{22}Gender_i + \beta_{32}PrevOI_i + \beta_{42}Stratum_i + \gamma_1 b_{0i} + \gamma_2 b_{1i}$$

Dalam pemodelan bersama, adanya hubungan antara hasil pengukuran longitudinal dan peubah respon primer yang menjadi perhatian ditunjukkan oleh hasil uji terhadap parameter γ_1 dan γ_2 dalam submodel-2. Pada Tabel 3 terlihat bahwa kedua parameter γ_1 dan γ_2 berbeda nyata dengan nol pada $\alpha = 5\%$, artinya baik jumlah sel CD4⁺ awal dari pasien (pada saat masuk dalam studi) maupun laju perubahannya per satuan waktu berpengaruh terhadap peluang kematian pasien. Nilai dugaan untuk γ_1 sebesar -0.2726 dengan selang kepercayaan 95% sebesar (-0.3624, -0.1828). Tanda negatif menunjukkan semakin banyak jumlah sel CD4⁺ awal dari pasien maka peluang

terjadinya kematian pada pasien semakin rendah, dan sebaliknya. Adapun nilai dugaan bagi parameter γ_2 sebesar -3.9857 dengan selang kepercayaan 95% sebesar (-7.2904, -0.6810), yang mengindikasikan bahwa semakin besar laju penurunan jumlah sel CD4⁺ pasien per bulan, maka peluang kematian pasien akan semakin tinggi.

Ragam intersep yang nyata pada $\alpha = 5\%$ dengan nilai dugaan ragam sebesar 15.8739 menunjukkan adanya keragaman jumlah sel CD4⁺ antar pasien sewaktu masuk dalam studi, dengan keragaman berkisar antara 13.5809 - 18.1670 pada taraf kepercayaan 95%. Ragam slope juga nyata dengan nilai dugaan sebesar 0.02913, artinya laju penurunan jumlah sel CD4⁺ per bulan bervariasi antar pasien dengan keragaman berkisar 0.01739 - 0.04086 pada taraf kepercayaan 95%. Terdapat korelasi negatif antara intersep dan slope, yang ditunjukkan oleh nilai peragam antara intersep dan slope sebesar -0.0918, atau korelasinya sebesar $r = -0.315$. Namun hasil pengujian tidak nyata pada $\alpha = 5\%$, yang berarti penurunan jumlah sel CD4⁺ antar pasien tidak dipengaruhi oleh jumlah sel CD4⁺ yang dimiliki sebelumnya (sewaktu masuk dalam studi).

Peubah bebas yang berpengaruh nyata pada $\alpha = 5\%$ terhadap respon primer terjadi atau tidak terjadinya kematian pasien pada submodel-2 adalah Drug dan PrevOI, sedangkan peubah Gender dan Stratum tidak nyata. Nilai dugaan untuk koefisien Drug sebesar 0.4914. Nilai positif ini menunjukkan bahwa peluang kematian pasien yang menggunakan obat ddI lebih tinggi dibandingkan pasien yang menerima obat ddC. Nilai dugaan parameter PrevOI sebesar 0.9807, yang mengindikasikan bahwa pasien yang terdiagnosis AIDS pada awal studi mempunyai peluang kematian yang lebih tinggi dibandingkan yang tidak terdiagnosis AIDS.

Nilai dugaan rasio odds untuk Drug sebesar $\exp(0.4914) = 1.635$ dengan selang kepercayaan 95% yaitu (1.026, 2.065), artinya dalam populasi kematian pasien terjadi antara 1.026 - 2.065 kali lebih sering di antara penerima obat ddI dibandingkan ddC. Demikian pula nilai dugaan rasio odds untuk PrevOI sebesar $\exp(0.9807) = 2.666$ dengan selang kepercayaan 95% yaitu (1.857, 3.828), dengan kata lain pasien yang terdiagnosis AIDS pada awal studi memiliki resiko kematian antara 1.857 - 3.828 kali dibandingkan pasien yang tidak terdiagnosis AIDS.

Tabel 3. Nilai dugaan parameter model bersama beserta hasil uji dan SK 95%

Parameter	Estimate	SE	tValue	Prob> t	95% CI	
Submodel-1						
Intercept (β_{01})	8.0164	0.3520	22.77	<.0001	7.3246	8.7082
Time (β_{11})	-0.1880	0.02163	-8.69	<.0001	-0.2305	-0.1455
Time x Drug (β_{21})	0.01981	0.02929	0.68	0.4991	-0.03774	0.07736
Gender (β_{31})	-0.1367	0.3263	-0.42	0.6756	-0.7779	0.5046
PrevOI (β_{41})	-2.3410	0.2397	-9.77	<.0001	-2.8119	-1.8700
Stratum (β_{51})	-0.1349	0.2365	-0.57	0.5686	-0.5996	0.3297
σ^2_{bo}	15.8739	1.1669	13.60	<.0001	13.5809	18.1670
$\sigma_{bo,b1}$	-0.09188	0.06193	-1.48	0.1386	-0.2136	0.02982
σ^2_{b1}	0.02913	0.005972	4.88	<.0001	0.01739	0.04086
σ^2	3.0501	0.1691	18.04	<.0001	2.7178	3.3823
Submodel-2						
Intercept (β_{02})	-1.0535	0.2829	-3.72	0.0002	-1.6094	-0.4975
Drug (β_{12})	0.4914	0.2377	2.07	0.0393	0.02421	0.9586
Gender (β_{22})	-0.1625	0.2182	-0.74	0.4569	-0.5913	0.2663
PrevOI (β_{32})	0.9807	0.1845	5.32	<.0001	0.6182	1.3432
Stratum (β_{42})	0.1996	0.1446	1.38	0.1680	-0.08448	0.4838
γ_1	-0.2726	0.04571	-5.96	<.0001	-0.3624	-0.1828
γ_2	-3.9857	1.6817	-2.37	0.0182	-7.2904	-0.6810

KESIMPULAN DAN SARAN

Pendekatan kekar dalam penelitian ini dicobakan terhadap series data longitudinal yang pendek. Dalam kondisi ini, pendekatan kekar atas dasar sebaran t untuk galat intra subyek dalam pemodelan bersama memberikan hasil yang lebih baik (ARB dan MSE yang lebih kecil), terutama untuk parameter penghubung submodel-1 dan submodel-2. Untuk sebaran galat yang berekor sangat panjang, disarankan menggunakan pendekatan sebaran t daripada sebaran normal untuk memperkecil bias dan memperkecil ragam dugaan, terutama untuk data berukuran kecil.

DAFTAR PUSTAKA

- Henderson R, Diggle P, Dobson A. 2000. Joint modelling of longitudinal measurement and event time data. *Biostatistics* 1(4):465-480.
- Guo X, Carlin BP. 2004. Separate and joint modeling of longitudinal and event time data using standard computer packages. *Am Statist* 58(1):1-9.
- Horrocks J, Heuvel MJ van Den. 2009. Prediction of pregnancy: a joint model for longitudinal and binary data. *Bayesian Analysis* 4(3):523-538.
- Kotz S, Nadarajah S. 2004. *Multivariate t Distributions and Their Applications*. New York: Cambridge University Press.
- Laird NM, Ware JH. 1982. Random effects models for longitudinal data. *Biometrics* 38:936-974.
- Lange KL, Little RJA, Taylor JMG. 1989. Robust statistical modeling using the t distribution. *J Amer Statist Assoc.* 84(408):881-896.
- Li E, Zhang D, Davidian M. 2004. Conditional estimation for generalized linear model when covariates are subject-specific parameters in a mixed model for longitudinal measurement. *Biometrics* 60:1-7.
- Li E, Wang N, Davidian M. 2007a. Joint models for a primary endpoint and multiple longitudinal covariate processes. *Biometrics* 63:1068-1078.
- Li E, Wang N, Davidian M. 2007b. Likelihood and pseudo-likelihood methods for semiparametric joint model for a primary endpoint and longitudinal data.

- Computational Statistics and Data Analysis* 51:5776-5790.
- Li N, Elashoff RM, Li G. 2009. Robust joint model of longitudinal measurements and competing risks failure time data. *Biometrics* 63(1):19-30.
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
- Nadarajah S, Kotz S. 2008. Estimation methods for the multivariate t distribution. *Acta Appl Math* 102:99-118.
- Pinheiro JC, Liu CH, Wu YN. 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics* 10:249-276.
- Song P, Zhang P, Qu A. 2007. Maximum likelihood inference in robust linear mixed-effects models using multivariate t distribution. *Statistica Sinica* 17:929-943.
- Tsiatis AA, Davidian M. 2004. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 14:809-834.
- Verbeke G, Molenberghs G. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Wang CY, Wang N, Wang S. 2000. Regression analysis when covariates are regression parameters of a random effects model. *Biometrics* 56:487-495.
- Wang WL, Fan TH. 2010. Estimation in multivariate t linear mixed models for multiple longitudinal data. Manuscript ID SS-09-306RI.
<http://www.stat.sinica.edu.tw/statistica/>
- Zellner, A. 1976. Bayesian and non-bayesian analysis of the regression model with multivariate student- t error terms. *J Amer Statist Assoc.* 71:400-405.
- Zhang D, Lin X. 1999. Generalized linear models with longitudinal covariates. Unpublished manuscript.