

Pendekatan Bayes Empirik Pada Pendugaan Statistik Area Kecil Berbasis Model Poisson-Gamma Dengan Peubah Penyerta

Kismiantini
Jurusan Pendidikan Matematika
FMIPA Universitas Negeri Yogyakarta

Abstrak

Bayes empirik merupakan salah satu metode pada pendugaan area kecil yang dapat digunakan untuk menduga parameter area kecil. Area kecil didefinisikan sebagai subpopulasi (area) yang berukuran contoh kecil. Metode Bayes empirik ini cocok digunakan pada data cacahan dengan model Poisson-Gamma selain metode Bayes hierarkhi. Tujuan penelitian ini adalah mengkaji penggunaan metode Bayes empirik pada pendugaan statistik area kecil berbasis model Poisson-Gamma dengan peubah penyerta. Hasil penelitian menunjukkan bahwa penduga statistik area kecil untuk data cacahan yakni penduga resiko relatif Bayes empirik dari model Poisson-Gamma dengan peubah penyerta memberikan hasil pendugaan dengan ketelitian yang lebih tinggi dibandingkan penduga langsung *standardized mortality ratio*.

Kata kunci: Bayes empirik, statistik area kecil, model Poisson-Gamma

PENDAHULUAN

Latar Belakang

Pendugaan area kecil (*small area estimation*) merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil. Teknik pendugaan ini memanfaatkan data dari domain besar (yakni seperti data sensus, data survei sosial ekonomi nasional) untuk menduga peubah yang menjadi perhatian pada domain yang lebih kecil. Area kecil didefinisikan sebagai subpopulasi yang ukuran contohnya kecil sehingga pendugaan langsung tidak dapat menghasilkan dugaan yang teliti (Rao 2003). Area kecil tersebut dapat berupa kota, kabupaten, kecamatan, desa/kelurahan, kelompok suku, kelompok jenis kelamin atau kelompok umur. Pendugaan langsung (*direct estimation*) pada subpopulasi relatif tidak memiliki presisi yang memadai karena kecilnya jumlah contoh yang digunakan untuk memperoleh dugaan tersebut.

Teknik pendugaan alternatif lain dikembangkan untuk meningkatkan keefektifan ukuran contoh dan menurunkan galat baku yakni pendugaan tak langsung (*indirect estimation*), pendugaan ini bersifat meminjam kekuatan dari pengamatan contoh area yang berdekatan dengan memanfaatkan informasi tambahan yakni dari data sensus dan catatan administratif (Rao 2005). Penduga yang dihasilkan pada pendugaan tak langsung ini adalah penduga prediksi tak bias linier terbaik empirik atau *empirical best linear unbiased prediction* selanjutnya disebut *EBLUP*, penduga Bayes empirik atau *empirical Bayes* disingkat *EB*, dan penduga Bayes hierarkhi atau *hierarchical Bayes* yang disingkat *HB*. Metode *EBLUP* merupakan metode yang diterapkan pada model

campuran linier, namun model campuran linier dirancang untuk peubah kontinu sehingga tidak cocok untuk data biner atau cacahan. Sehingga untuk data biner atau cacahan digunakan metode *EB* dan *HB* untuk melakukan pendugaan area kecil.

Salah satu penerapan pendugaan area kecil untuk data cacahan adalah pada pemetaan penyakit. Pada pemetaan penyakit, ukuran contoh (jumlah kasus berpenyakit) yang kecil merupakan suatu masalah yang sering dihadapi karena areanya yang sangat kecil, penyakit jarang terjadi atau keduanya. Sehingga pendugaan langsung dalam menduga resiko relatif yaitu *standardized mortality ratio (SMR)* menjadi tidak dapat diandalkan. Metode alternatif untuk menangani permasalahan tersebut adalah metode Bayes empirik, dengan model yang sering digunakan adalah model Poisson-Gamma. Kelebihan dari metode Bayes empirik ini diantaranya dikemukakan oleh Marshall (1991) yakni mampu menampung informasi antar area yang dimaksudkan mereduksi jumlah kuadrat tengah galat dan Pringle (1995) yang menyatakan bahwa teknik Bayes empirik cocok digunakan karena menghasilkan penduga resiko relatif yang lebih andal dibandingkan penduga kemungkinan maksimum. Upaya perbaikan pendugaan resiko relatif dapat dilakukan dengan menggunakan model Poisson-Gamma dengan peubah penyerta.

Rumusan Masalah

Bagaimana penggunaan metode Bayes empirik pada pendugaan statistik area kecil berbasis model Poisson-Gamma dengan peubah penyerta?

Tujuan Penelitian

Penelitian ini bertujuan untuk mengkaji penggunaan metode Bayes empirik pada pendugaan statistik area kecil berbasis model Poisson-Gamma dengan peubah penyerta.

Manfaat Penelitian

Penelitian ini diharapkan dapat menambah wawasan pembaca mengenai pendugaan area kecil untuk data cacahan.

TINJAUAN PUSTAKA

Model Area Kecil

Model area kecil merupakan model dasar dalam pendugaan area kecil. Model ini dikelompokkan menjadi dua yakni model level area dasar dan model level unit dasar (Rao 2003). Pada model level area dasar diasumsikan bahwa peubah yang menjadi perhatian merupakan fungsi dari rata-rata peubah respon, $\theta_i = g(\bar{Y}_i)$ untuk $g(\cdot)$ tertentu, yang berkaitan dengan data penyerta area kecil $\underline{z}_i = (z_{i1}, \dots, z_{ip})^T$ dan mengikuti model linier sebagai berikut :

$$\theta_i = \underline{z}_i^T \underline{\beta} + b_i v_i, \quad i = 1, \dots, m \quad (1)$$

dengan b_i adalah konstanta bernilai positif yang diketahui dan $\underline{\beta} = (\beta_1, \dots, \beta_p)^T$ adalah vektor koefisien regresi berukuran $p \times 1$. Sedangkan v_i adalah pengaruh acak area kecil yang diasumsikan memiliki sebaran identik dan saling bebas yakni

$$E(v_i) = 0, \quad \text{Var}(v_i) = \sigma_v^2 (\geq 0) \quad (2)$$

Seringkali pengaruh acak v_i dianggap menyebar normal.

Untuk melakukan inferensi tentang rata-rata area kecil \bar{Y}_i pada model (1), dimisalkan bahwa penduga langsung \hat{Y}_i ada dan

$$\hat{\theta}_i = g(\hat{Y}_i) = \theta_i + e_i, \quad i = 1, \dots, m \quad (3)$$

dengan galat contoh e_i bebas serta

$$E(e_i | \theta_i) = 0, \quad \text{Var}(e_i | \theta_i) = \psi_i \quad (4)$$

Biasanya ragam contoh ψ_i diketahui.

Dengan menggabungkan persamaan (1) dan (3) maka diperoleh model

$$\hat{\theta}_i = \underline{z}_i^T \underline{\beta} + b_i v_i + e_i, \quad i = 1, \dots, m \quad (5)$$

dengan v_i dan e_i saling bebas. Model (5) ini merupakan kasus khusus dari model campuran linier dan dikenal pula sebagai model Fay-Herriot dalam literatur area kecil (Rao 2003). Fay dan Herriot (1979) menggunakan model level area dasar tersebut untuk

menduga pendapatan per kapita bagi beberapa area kecil di Amerika Serikat dengan populasi kurang dari 1000.

Sedangkan untuk model level unit dasar mengasumsikan bahwa data penyerta unit $\underline{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ ada untuk masing-masing anggota populasi j dalam masing-masing area kecil i , namun kadang cukup dengan rata-rata populasi \bar{x}_i diketahui saja. Selanjutnya peubah perhatian y_{ij} dianggap berkaitan dengan \underline{x}_{ij} mengikuti model regresi linier galat tersarang satu tahap sebagai berikut :

$$y_{ij} = \underline{x}_{ij}^T \underline{\beta} + v_i + e_{ij}; \quad j = 1, \dots, N, i = 1, \dots, m \quad (6)$$

Pengaruh acak area kecil v_i ini mempunyai sebaran identik dan saling bebas, $e_{ij} = k_{ij} \tilde{e}_{ij}$ dengan konstanta k_{ij} diketahui dan \tilde{e}_{ij} peubah acak yang memiliki sebaran identik dan bebas pula serta bebas dengan v_i ,

$$E(\tilde{e}_{ij}) = 0, \quad \text{Var}(\tilde{e}_{ij}) = \sigma_e^2 \quad (7)$$

Seringkali diasumsikan bahwa v_i dan e_i berdistribusi normal.

Model level area digunakan pula dalam pemetaan penyakit (Ghosh & Rao 1994) yakni dengan menganggap bahwa banyaknya pengamatan suatu kasus dari area kecil $y_i \sim \overset{ind}{Poisson}(e_i \theta_i)$ dan $\theta_i \sim \overset{iid}{gamma}(v, \alpha)$, θ_i menyatakan resiko relatif area ke- i yang tidak diketahui dan e_i adalah nilai harapan banyaknya suatu kasus pada area ke- i .

Penduga sederhana resiko relatif pada pemetaan penyakit adalah *standardized mortality ratio* atau disingkat sebagai *SMR* (Wakefield & Elliott 1999), yang selanjutnya disebut sebagai penduga langsung dalam pendugaan area kecil. *SMR* merupakan penduga statistik resiko relatif yang berguna untuk mengetahui sebaran geografis penyakit. Penduga ini diperoleh dari asumsi umum pemetaan penyakit bahwa banyaknya pengamatan suatu kasus yaitu $y_i \sim \overset{ind}{Poisson}(e_i \theta_i)$ dengan fungsi peluang adalah

$$f(y_i | \theta_i) = \frac{e^{-e_i \theta_i} (e_i \theta_i)^{y_i}}{y_i!}, \quad y_i = 0, 1, \dots; i = 1, \dots, m \quad (8)$$

Fungsi kemungkinan adalah

$$L(\theta_i | y_i) = f(y_i | \theta_i) = \frac{e^{-e_i \theta_i} (e_i \theta_i)^{y_i}}{y_i!} \quad (9)$$

Logaritma natural dari fungsi kemungkinan adalah

$$\ln L(\theta_i | y_i) = -e_i \theta_i + y_i \ln(e_i \theta_i) - \ln y_i! \quad (10)$$

Sehingga

$$\begin{aligned} \frac{d}{d\theta_i} \ln L(\theta_i | y_i) &= -e_i + \frac{y_i e_i}{e_i \theta_i} = 0 \\ \Rightarrow \hat{\theta}_i &= \frac{y_i}{e_i} \end{aligned} \quad (11)$$

$\hat{\theta}_i = y_i/e_i$ merupakan penduga tak bias, yang ditunjukkan dengan harapan penduga parameter sama dengan parameternya.

$$E(\hat{\theta}_i) = E\left(\frac{Y_i}{e_i}\right) = \frac{1}{e_i} E(Y_i) = \frac{1}{e_i} e_i \theta_i = \theta_i \quad (12)$$

dengan

$$\begin{aligned} E(Y_i) &= \sum_{y_i=0}^{\infty} y_i \frac{e^{-e_i \theta_i} (e_i \theta_i)^{y_i}}{y_i!} = 0 + \sum_{y_i=1}^{\infty} y_i \frac{e^{-e_i \theta_i} (e_i \theta_i)^{y_i}}{y_i!} = e^{-e_i \theta_i} e_i \theta_i \sum_{y_i=1}^{\infty} \frac{(e_i \theta_i)^{y_i-1}}{(y_i-1)!} \\ &= e^{-e_i \theta_i} e_i \theta_i e^{e_i \theta_i} = e_i \theta_i \end{aligned} \quad (13)$$

Penduga bagi θ_i yakni $\hat{\theta}_i = y_i/e_i$ ini merupakan penduga kemungkinan maksimum yang bersifat tak bias.

Metode Bayes Empirik

Metode Bayes sederhana yaitu misalkan kemungkinan $Y|\theta \sim f(y|\theta)$ dan prior $\theta \sim \pi(\theta)$, Y atau θ dapat berupa vektor dan π diasumsikan diketahui, maka sebaran posterior dari θ adalah

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{m(y)}, \text{ dengan } m(y) = \int f(y|\theta)\pi(\theta)d\theta. \quad (14)$$

Bila informasi parameter pada sebaran awal tidak diketahui, lalu dengan memanfaatkan data diperoleh nilai dugaan parameternya maka metode Bayes ini disebut sebagai Bayes empirik.

Metode Bayes empirik merupakan metode yang cocok digunakan dalam menangani data biner dan data cacahan pada pendugaan area kecil. Metode Bayes empirik dalam konteks pendugaan area kecil secara ringkas adalah (Rao 2003):

1. mendapatkan fungsi kepekatan peluang akhir (posterior) dari parameter area kecil yang menjadi perhatian
2. menduga parameter model dari fungsi kepekatan peluang marjinal
3. menggunakan fungsi kepekatan peluang posterior dugaan untuk membuat inferensi parameter area kecil yang menjadi perhatian.

Model Poisson-Gamma dengan Peubah penyerta

Wakefield (2006) memperkenalkan model Poisson-Gamma dengan peubah penyerta yang secara analitik lebih mudah dikerjakan dan dalam pendugaan. Pada tahap pertama, diasumsikan bahwa $y_i \sim \text{Poisson}(e_i \mu_i \theta_i)$ dengan $\mu_i = \mu(\underline{x}_i, \underline{\beta})$ menyatakan model regresi sehingga $\underline{x}_i = (x_{i1}, \dots, x_{pi})^T$ merupakan vektor peubah penyerta tetap dan $\underline{\beta} = (\beta_1, \dots, \beta_p)^T$ merupakan vektor koefisien regresi. Tahap kedua diasumsikan bahwa $\theta_i \sim \text{gamma}(\alpha, \alpha)$ yang selanjutnya sebagai prior dengan rata-rata 1 dan ragam $1/\alpha$.

Fungsi peluang dari $y_i \sim \text{Poisson}(e_i \mu_i \theta_i)$ adalah

$$f(y_i | \theta_i) = \frac{e^{-e_i \mu_i \theta_i} (e_i \mu_i \theta_i)^{y_i}}{y_i!}, \quad y_i = 0, 1, \dots \quad (15)$$

Fungsi kepadatan peluang dari $\theta_i \sim \text{gamma}(\alpha, \alpha)$ adalah

$$\pi(\theta_i) = \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha \theta_i} \theta_i^{\alpha-1}, \quad \theta_i > 0 \quad (16)$$

Sehingga fungsi bersamanya diperoleh

$$f(y_i, \theta_i) = \frac{e^{-e_i \mu_i \theta_i} (e_i \mu_i \theta_i)^{y_i}}{y_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha \theta_i} \theta_i^{\alpha-1}, \quad y_i = 0, 1, \dots; \theta_i > 0 \quad (17)$$

Selanjutnya fungsi marjinalnya adalah

$$\begin{aligned}
 m(y_i) &= \int_0^{\infty} f(y_i, \theta_i) d\theta_i \\
 &= \int_0^{\infty} \frac{e^{-e_i \mu_i \theta_i} (e_i \mu_i \theta_i)^{y_i}}{y_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha \theta_i} \theta_i^{\alpha-1} d\theta_i \\
 &= \frac{\alpha^\alpha (e_i \mu_i)^{y_i}}{\Gamma(\alpha) y_i!} \int_0^{\infty} e^{-e_i \mu_i \theta_i} \theta_i^{y_i} e^{-\alpha \theta_i} \theta_i^{\alpha-1} d\theta_i \\
 &= \frac{\alpha^\alpha (e_i \mu_i)^{y_i}}{\Gamma(\alpha) y_i!} \int_0^{\infty} e^{-(e_i \mu_i + \alpha) \theta_i} \theta_i^{y_i + \alpha - 1} d\theta_i \\
 &= \frac{\alpha^\alpha (e_i \mu_i)^{y_i}}{\Gamma(\alpha) y_i!} \int_0^{\infty} e^{-t_i} \left(\frac{t_i}{e_i \mu_i + \alpha} \right)^{y_i + \alpha - 1} \frac{1}{e_i \mu_i + \alpha} dt_i \\
 &= \frac{\alpha^\alpha (e_i \mu_i)^{y_i}}{\Gamma(\alpha) y_i! (e_i \mu_i + \alpha)^{y_i + \alpha}} \int_0^{\infty} e^{-t_i} t_i^{y_i + \alpha - 1} dt_i \\
 &= \frac{\alpha^\alpha (e_i \mu_i)^{y_i}}{\Gamma(\alpha) y_i! (e_i \mu_i + \alpha)^{y_i + \alpha}} \Gamma(y_i + \alpha) \\
 &= \binom{y_i + \alpha - 1}{\alpha - 1} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^\alpha \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i} \tag{18}
 \end{aligned}$$

misal : $t_i = (e_i \mu_i + \alpha) \theta_i$
 $dt_i = (e_i \mu_i + \alpha) d\theta_i$
 batas :
 $\theta_i = 0 \rightarrow t_i = 0$
 $\theta_i = \infty \rightarrow t_i = \infty$

Fungsi marjinal diatas merupakan fungsi sebaran binomial negatif dengan rata-rata dan ragam berikut :

$$\begin{aligned}
 E[Y_i | \underline{\beta}, \alpha] &= e_i \mu_i, \quad Var(Y_i | \underline{\beta}, \alpha) = E[Y_i | \underline{\beta}, \alpha] \left(1 + E[Y_i | \underline{\beta}, \alpha] / \alpha \right) = e_i \mu_i (1 + e_i \mu_i / \alpha) \\
 &\tag{19}
 \end{aligned}$$

Sehingga ragam meningkat sebagai fungsi kuadratik dari rata-rata, dan parameter skala α dapat mengakomodasi overdispersi. Berikut cara memperoleh rata-rata dan ragam dari sebaran binomial negatif.

$$\begin{aligned}
 E(Y_i) &= \sum_{y_i=0}^{\infty} y_i \binom{y_i + \alpha - 1}{\alpha - 1} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^\alpha \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i} \\
 &= \sum_{y_i=1}^{\infty} y_i \frac{(y_i + \alpha - 1)!}{y_i! (\alpha - 1)!} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^\alpha \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i} \\
 &= \alpha \left(\frac{e_i \mu_i + \alpha}{\alpha} \right) \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right) \sum_{y_i=1}^{\infty} \frac{(y_i + \alpha - 1)!}{(y_i - 1)! \alpha!} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^{\alpha+1} \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i-1} \\
 &= e_i \mu_i \sum_{y_i=1}^{\infty} \binom{y_i + \alpha - 1}{y_i - 1} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^{\alpha+1} \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i-1} \\
 &= e_i \mu_i \left[\sum_{z_i=0}^{\infty} \binom{(\alpha + 1) + z_i - 1}{z_i} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^{\alpha+1} \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{z_i} \right] \\
 &= e_i \mu \times 1 \\
 &= e_i \mu_i
 \end{aligned}$$

Misal :
 $z_i = y_i - 1$
 $y_i = 1 \rightarrow z_i = 0$
 $y_i = \infty \rightarrow z_i = \infty$

$$\begin{aligned}
 E(Y_i(Y_i - 1)) &= \sum_{y_i=0}^{\infty} y_i(y_i - 1) \binom{y_i + \alpha - 1}{\alpha - 1} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^\alpha \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i} \\
 &= \sum_{y_i=2}^{\infty} y_i(y_i - 1) \frac{(y_i + \alpha - 1)!}{y_i!(\alpha - 1)!} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^\alpha \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i} \\
 &= (\alpha + 1) \alpha \left(\frac{e_i \mu_i + \alpha}{\alpha} \right)^2 \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^2 \sum_{y_i=2}^{\infty} \frac{(y_i + \alpha - 1)!}{(y_i - 2)!(\alpha + 1)!} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^{\alpha+2} \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i-2} \\
 &= \frac{(\alpha + 1)}{\alpha} (e_i \mu_i)^2 \sum_{y_i=2}^{\infty} \binom{y_i + \alpha - 1}{y_i - 2} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^{\alpha+2} \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{y_i-2} \\
 &= \frac{(\alpha + 1)}{\alpha} (e_i \mu_i)^2 \left[\sum_{z_i=0}^{\infty} \binom{(\alpha + 2) + z_i - 1}{z_i} \left(\frac{\alpha}{e_i \mu_i + \alpha} \right)^{\alpha+2} \left(1 - \frac{\alpha}{e_i \mu_i + \alpha} \right)^{z_i} \right] \\
 &= \frac{(\alpha + 1)}{\alpha} (e_i \mu_i)^2 \times 1 \\
 &= \frac{(\alpha + 1)}{\alpha} (e_i \mu_i)^2
 \end{aligned}$$

Misal :
 $z_i = y_i - 2$
 $y_i = 2 \rightarrow z_i = 0$
 $y_i = \infty \rightarrow z_i = \infty$

$E(Y_i(Y_i - 1)) = E(Y_i^2) - E(Y_i) \Rightarrow E(Y_i^2) = \frac{(\alpha + 1)}{\alpha} (e_i \mu_i)^2 + (e_i \mu_i)$. Sehingga

$Var(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = \frac{(\alpha + 1)}{\alpha} (e_i \mu_i)^2 + e_i \mu_i - (e_i \mu_i)^2 = e_i \mu_i \left(1 + \frac{e_i \mu_i}{\alpha} \right)$.

Selanjutnya dugaan parameter prior, yaitu $\hat{\beta}$ dan $\hat{\alpha}$, dapat diperoleh dari sebaran marjinal $y_i | \beta, \alpha \stackrel{iid}{\sim}$ **binomial negatif** menggunakan pendugaan kemungkinan maksimum, yang merupakan penyelesaian dari teknik regresi binomial negatif. Berdasarkan teorema Bayes maka fungsi posterior adalah sebagai berikut :

$$\begin{aligned}
 \pi(\theta_i | y_i, \beta, \alpha) &= \frac{f(y_i, \theta_i)}{m(y_i)} = \frac{\frac{e^{-e_i \mu_i \theta_i} (e_i \mu_i \theta_i)^{y_i}}{y_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha \theta_i} \theta_i^{\alpha-1}}{\frac{\alpha^\alpha (e_i \mu_i)^{y_i}}{\Gamma(\alpha) y_i! (e_i \mu_i + \alpha)^{y_i + \alpha}} \Gamma(y_i + \alpha)} \quad (20) \\
 &= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} e^{-(e_i \mu_i + \alpha) \theta_i} \theta_i^{y_i + \alpha - 1}, \theta_i > 0
 \end{aligned}$$

Sehingga $\theta_i | y_i, \beta, \alpha \sim \text{gamma}(y_i + \alpha, e_i \mu_i + \alpha)$. Dari fungsi posterior tersebut diperoleh penduga Bayes bagi θ_i dan ragam posterior bagi θ_i adalah

$\hat{\theta}_i^B(\beta, \alpha) = E(\theta_i | y_i, \beta, \alpha) = (y_i + \alpha) / (e_i \mu_i + \alpha)$ dan

$$V(\theta_i | y_i, \beta, \alpha) = g_{1i}(\beta, \alpha, y_i) = (y_i + \alpha) / (e_i \mu_i + \alpha)^2 \quad (21)$$

Penduga Bayes bagi θ_i merupakan nilai harapan dari posterior berikut :

$$\begin{aligned} E(\theta_i | y_i, \beta, \alpha) &= \int_0^\infty \theta_i \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} e^{-(e_i \mu_i + \alpha)\theta_i} \theta_i^{y_i + \alpha - 1} d\theta_i \\ &= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \int_0^\infty e^{-(e_i \mu_i + \alpha)\theta_i} \theta_i^{y_i + \alpha - 1} d\theta_i \\ &= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \int_0^\infty e^{-t_i} \left(\frac{t_i}{e_i \mu_i + \alpha} \right)^{y_i + \alpha - 1} \frac{1}{(e_i \mu_i + \alpha)} dt_i \\ &= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \frac{1}{(e_i \mu_i + \alpha)^{y_i + \alpha}} \int_0^\infty e^{-t_i} (t_i)^{y_i + \alpha - 1} dt_i \\ &= \frac{1}{(e_i \mu_i + \alpha) \Gamma(y_i + \alpha)} \Gamma(y_i + \alpha) \\ &= \frac{(y_i + \alpha) \Gamma(y_i + \alpha)}{(e_i \mu_i + \alpha) \Gamma(y_i + \alpha)} \\ &= \frac{(y_i + \alpha)}{(e_i \mu_i + \alpha)} \end{aligned}$$

misal : $t_i = (e_i \mu_i + \alpha)\theta_i$ $dt_i = (e_i \mu_i + \alpha)d\theta_i$ batas : $\theta_i = 0 \rightarrow t_i = 0$ $\theta_i = \infty \rightarrow t_i = \infty$

Sedangkan untuk mendapatkan ragam posterior bagi θ_i , dicari terlebih dahulu

$$E(\theta_i^2 | y_i, \beta, \alpha).$$

$$\begin{aligned}
E(\theta_i^2 | y_i, \beta, \alpha) &= \int_0^\infty \theta_i^2 \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} e^{-(e_i + \alpha)\theta_i} \theta_i^{y_i + \alpha - 1} d\theta_i \\
&= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \int_0^\infty e^{-(e_i \mu_i + \alpha)\theta_i} \theta_i^{y_i + \alpha + 2 - 1} d\theta_i \\
&= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \int_0^\infty e^{-t_i} \left(\frac{t_i}{e_i \mu_i + \alpha} \right)^{y_i + \alpha + 2 - 1} \frac{1}{(e_i \mu_i + \alpha)} dt_i \\
&= \frac{(e_i \mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \frac{1}{(e_i + \alpha)^{y_i + \alpha + 2}} \int_0^\infty e^{-t_i} (t_i)^{y_i + \alpha + 2 - 1} dt_i \\
&= \frac{1}{(e_i \mu_i + \alpha)^2 \Gamma(y_i + \alpha)} \Gamma(y_i + \alpha + 2) \\
&= \frac{(y_i + \alpha + 1)(y_i + \alpha) \Gamma(y_i + \alpha)}{(e_i + \alpha)^2 \Gamma(y_i + \alpha)} \\
&= \frac{(y_i + \alpha + 1)(y_i + \alpha)}{(e_i + \alpha)^2}
\end{aligned}$$

Sehingga ragam posterior bagi θ_i adalah

$$\begin{aligned}
V(\theta_i | y_i, \beta, \alpha) &= E(\theta_i^2 | y_i, \beta, \alpha) - [E(\theta_i | y_i, \beta, \alpha)]^2 \\
&= \frac{(y_i + \alpha + 1)(y_i + \alpha)}{(e_i \mu_i + \alpha)^2} - \left[\frac{(y_i + \alpha)}{(e_i \mu_i + \alpha)} \right]^2 \\
&= \frac{(y_i + \alpha)(y_i + \alpha + 1 - (y_i + \alpha))}{(e_i \mu_i + \alpha)^2} \\
&= \frac{(y_i + \alpha)}{(e_i \mu_i + \alpha)^2}
\end{aligned}$$

Selanjutnya penduga Bayes empirik bagi θ_i menurut Wakefield (2006) adalah sebagai berikut :

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^B(\hat{\beta}, \hat{\alpha}) = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) E[RR_i] \quad (22)$$

dengan $\hat{\gamma}_i = e_i \hat{\mu}_i / (\hat{\alpha} + e_i \hat{\mu}_i)$, $E[RR_i] = \hat{\mu}_i \times E[\theta_i] = \hat{\mu}_i \times 1 = \hat{\mu}_i = \exp(\underline{x}_i^T \hat{\beta})$, $E[RR_i]$ adalah nilai harapan resiko relatif ke- i yang merupakan penduga tak langsung, $\hat{\theta}_i = y_i / e_i$ adalah penduga langsung (*standardized mortality ratio*) dari θ_i , y_i dan e_i masing-masing menyatakan banyaknya pengamatan dan nilai harapan banyaknya suatu kasus.

Keoptimuman hubungan antara peubah penyerta dengan peubah yang menjadi perhatian dilihat dari ukuran kelayakan model regresi binomial negatif (*goodness of fit*) dan pemeriksaan sisaan secara grafis. Sisaan pada model regresi binomial negatif ini berbeda dengan sisaan pada regresi linier biasa. Model regresi binomial negatif merupakan bagian dari model linier terampat. Sisaan pada model linier terampat adalah sisaan Pearson dan sisaan *deviance* yang dirumuskan sebagai berikut (Agresti 2002) :

$$r_p = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(Y_i)}} \text{ dan } r_{D_i} = \sqrt{d_i} \times \text{sign}(y_i - \hat{\mu}_i) \quad (23)$$

dengan $d_i = 2\omega_i \left[y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right]$ adalah *deviance*.

Sehingga pemeriksaan secara grafis dari model regresi binomial negatif tersebut adalah dengan membuat plot antara prediktor linier dengan sisaan Pearson atau sisaan *deviance*.

DATA DAN METODE

Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari literatur. Data tersebut diambil dari Stern dan Cressie (2000) yakni tentang kanker bibir di Skotlandia. Data ini berupa banyaknya penderita kanker bibir (sebagai peubah respon) dan nilai harapan banyaknya penderita kanker bibir yang tercatat selama 6 tahun dari 1975 sampai 1980 pada masing-masing 56 distrik (area kecil) di Skotlandia. Sebagai peubah penyerta adalah persentase bekerja pada bidang pertanian, perikanan dan kehutanan.

Metode Penelitian

Metode penelitian yang digunakan dalam menduga resiko relatif suatu area terjangkit penyakit adalah berdasarkan penduga langsung *standardized mortality ratio* (SMR) dan penduga Bayes empirik dari model Poisson-Gamma dengan peubah penyerta yang diuraikan sebagai berikut :

A. Penduga langsung berdasarkan *standardized mortality ratio* (SMR) adalah

1. Menentukan e_i yaitu nilai harapan banyaknya suatu kasus pada subpopulasi ke- i dengan rumus

$$e_i = n_i \left(\frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m n_i} \right)$$

dengan y_i menyatakan banyaknya pengamatan suatu kasus pada subpopulasi ke- i , n_i menyatakan banyaknya individu pada subpopulasi ke- i dan m menyatakan jumlah subpopulasi. Subpopulasi ini dapat berupa distrik. Pada data kanker bibir, nilai harapan banyaknya suatu kasus pada subpopulasi ke- i sudah diketahui.

2. Menentukan *SMR* yaitu $\hat{\theta}_i = y_i/e_i$.
3. Menentukan dugaan kuadrat tengah galat (*KTG*) yaitu $\text{ktg}(\hat{\theta}_i) = (y_i/e_i)/e_i = y_i/e_i^2$. Diperoleh demikian karena penduga langsung $\hat{\theta}_i = y_i/e_i$ bersifat tak bias sehingga $\text{KTG}(\hat{\theta}_i) = \text{Var}(\hat{\theta}_i) = \theta_i/e_i$.
4. Proses hitungan dilakukan dengan Microsoft Office Excel.

B. Penduga Bayes empirik berdasarkan model Poisson-Gamma dengan peubah penyerta

1. Menentukan $\hat{\beta}$ dan $\hat{\alpha}$ dengan PROC GENMOD pada SAS versi 9.1.
2. Menentukan penduga Bayes empirik berdasarkan

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^B(\hat{\beta}, \hat{\alpha}) = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) e^{x_i^T \hat{\beta}}$$

$$\text{dengan } \hat{\gamma}_i = \left(e_i e^{x_i^T \hat{\beta}} \right) / \left(e_i e^{x_i^T \hat{\beta}} + \hat{\alpha} \right) \text{ dan } \hat{\theta}_i = y_i/e_i$$

3. Menentukan kuadrat tengah galat dengan menggunakan metode Jackknife yaitu :
 - Anggap bahwa $\hat{\theta}_i^{EB} = k_i(y_i, \hat{\beta}, \hat{\alpha})$, $\hat{\theta}_{i,-l}^{EB} = k_i(y_i, \hat{\beta}_{-l}, \hat{\alpha}_{-l})$, lalu

$$\hat{M}_{2i} = \frac{m-1}{m} \sum_{l=1}^m \left(\hat{\theta}_i^{EB} - \hat{\theta}_{i,-l}^{EB} \right)^2$$

- Dengan mencari $\hat{\beta}_{-l}$ dan $\hat{\alpha}_{-l}$ yang merupakan penduga kemungkinan maksimum yang diperoleh dari data ke- l yang dihapus, maka dihitung

$$\hat{M}_{1i} = g_{1i}(\hat{\beta}, \hat{\alpha}, y_i) - \frac{m-1}{m} \sum_{l=1}^m \left[g_{1i}(\hat{\beta}_{-l}, \hat{\alpha}_{-l}, y_i) - g_{1i}(\hat{\beta}, \hat{\alpha}, y_i) \right]$$

- Penduga Jackknife bagi kuadrat tengah galat penduga Bayes empirik diberikan oleh

$$\text{ktg}_J(\hat{\theta}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i}$$

4. Proses hitungan pada langkah 2 sampai 3 dilakukan dengan SAS/IML versi 9.1 dan Microsoft Office Excel. Pada penelitian ini digunakan perangkat lunak SAS versi 9.1 dengan PROC GENMOD dan IML untuk memperoleh dugaan parameter prior dan penduga Bayes empirik.

C. Pemeriksaan sisaan pada regresi binomial negatif

1. Menentukan sisaan Pearson.
2. Membuat plot antara prediktor linier dengan sisaan Pearson.
3. Proses 1 sampai 2 dilakukan dengan menggunakan PROC GENMOD pada SAS versi 9.1.

Perbandingan kebaikan antara *SMR* dan penduga Bayes empirik dari model Poisson-Gamma dengan peubah penyerta dengan melihat nilai galat baku.

HASIL PENELITIAN DAN PEMBAHASAN

Deskripsi Data

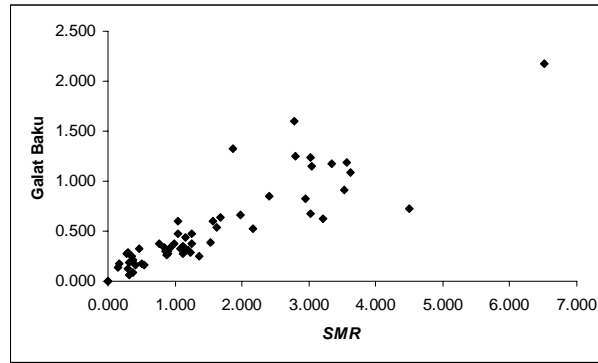
Tabel 1 menyajikan sebaran dari data kanker bibir pada 56 distrik di Skotlandia (Stern & Cressie 2000). Berdasarkan Tabel 1 dapat diketahui bahwa ada distrik yang tidak terjangkau penyakit, yang dinyatakan oleh nilai minimum penduga langsung *standardized mortality ratio (SMR)*. Padahal dalam kenyataannya bisa saja distrik tersebut mempunyai resiko relatif terjangkau penyakit. Tingginya nilai harapan banyaknya penderita kanker bibir yakni sebesar 88.608 merupakan akibat dari besarnya jumlah penduduk pada distrik tersebut. Sedangkan berdasarkan nilai simpangan baku, nilai harapan banyaknya penderita kanker bibir memiliki keragaman yang lebih besar dibandingkan banyaknya penderita kanker bibir (pengamatan).

Tabel 1 Statistik deskriptif dari data kanker bibir di Skotlandia

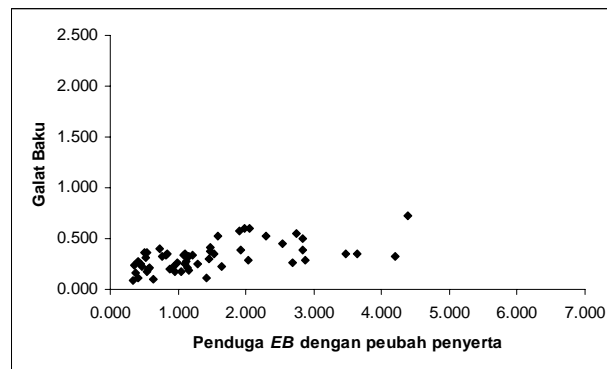
	Minimum	Maksimum	Rata-rata	Simpangan baku
Pengamatan	0	39	9.571	7.908
Harapan	1.070	88.608	9.571	13.164
<i>SMR</i>	0	6.522	1.526	1.315

Pendugaan Resiko Relatif

Pada penelitian ini, untuk mengetahui dugaan resiko relatif suatu distrik terjangkau penyakit kanker bibir yaitu dengan penduga langsung *SMR* dan penduga Bayes empirik dari model Poisson-Gamma dengan peubah penyerta. Kebaikan pendugaan resiko relatif diukur dari tingkat ketelitian yang ditunjukkan oleh besarnya galat baku. Grafis hubungan antara penduga resiko relatif dengan galat bakunya disajikan pada Gambar 1 dan 2.



Gambar 1 Plot *SMR* dengan galat baku.



Gambar 2 Plot penduga Bayes Empirik dengan peubah penyerta dan galat baku.

Gambar 1 dan 2 secara umum menunjukkan bahwa semakin besar nilai penduga resiko relatif semakin besar galat bakunya. Penduga Bayes empirik dengan peubah penyerta memberikan ketelitian yang lebih baik dibanding *SMR*. Hal ini terlihat bahwa untuk penduga Bayes empirik dengan peubah penyerta, semakin besar nilai penduga resiko relatif semakin kecil galat bakunya jika dibandingkan dengan penduga langsung *SMR*.

Tabel 2 Pendugaan data kanker bibir di Skotlandia

Rata-rata	<i>SMR</i>	Penduga Bayes empirik dengan peubah penyerta
Resiko Relatif	1.526	1.449
Kuadrat Tengah Galat	0.465	0.120
Galat Baku	0.527	0.318

Pada Tabel 2 diperoleh informasi bahwa secara rata-rata penduga Bayes empirik dari model Poisson-Gamma dengan peubah penyerta memberikan ketelitian yang lebih baik yaitu dengan nilai galat baku yang lebih kecil dibanding penduga langsung *SMR*.

Hal ini disebabkan oleh peubah penyerta yang dapat dimodelkan secara optimum dengan peubah yang menjadi perhatian. Hubungan yang optimum tersebut dijelaskan oleh nilai *goodness of fit* dan sisaan yang terdapat pada Tabel 3, Gambar 3 dan Gambar 4 berikut.

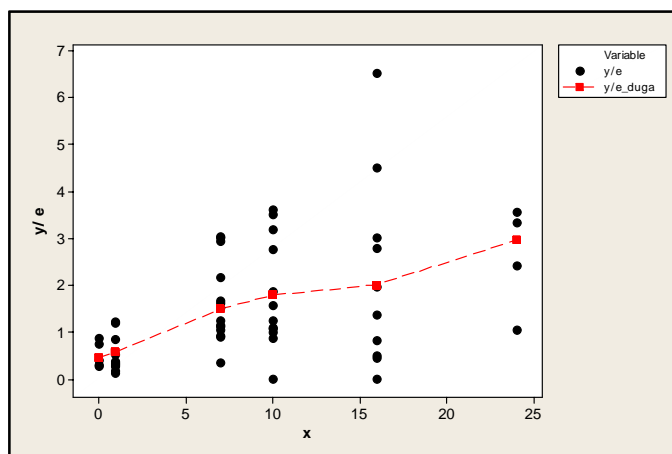
Tabel 3 Kriteria untuk mengukur kelayakan model

Criterion	DF	Value	Value/DF
Deviance	54	62.294	1.154
Scaled Deviance	54	62.294	1.154
Pearson Chi-Square	54	57.497	1.065
Scaled Pearson X2	54	57.497	1.065
Log Likelihood		770.728	

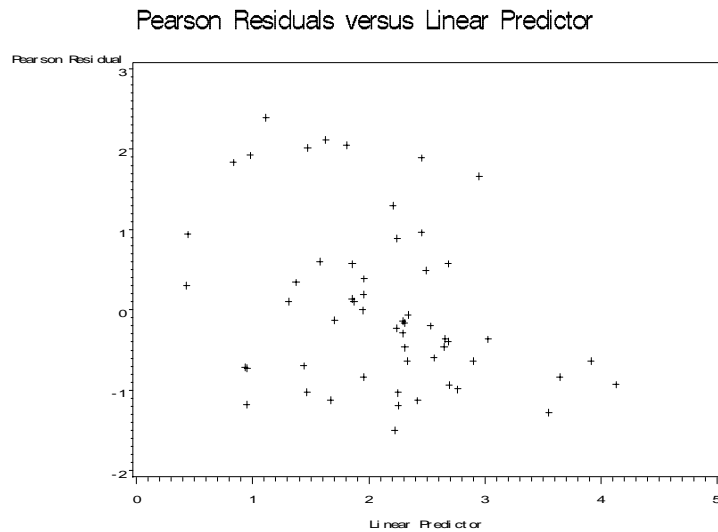
Ukuran kelayakan model di atas diperoleh dari teknik regresi binomial negatif, dengan persamaan regresi adalah :

$$\ln \left(\frac{\hat{y}_i}{e_i} \right) = -0.352 + 0.072x_i$$

Dari Tabel 3 tampak bahwa model regresi binomial negatif ini layak digunakan, karena nilai *Value/DF* untuk masing-masing kriteria besarnya kurang dari dua. Sedangkan Gambar 3 dan 4 juga menunjukkan terpenuhinya kelayakan model dengan persamaan regresi mencerminkan sebaran dari data dan sisaan cenderung tidak berpola.



Gambar 3 Hubungan antara resiko relatif (y/e) dengan persentase bekerja pada bidang pertanian, perikanan dan kehutanan (x).



Gambar 4 Prediktor linier dengan sisaan Pearson.

SIMPULAN

Beberapa simpulan yang diperoleh dari hasil penelitian ini adalah :

1. Penduga resiko relatif Bayes empirik dari model Poisson-Gamma dengan peubah penyerta memberikan hasil pendugaan dengan ketelitian yang lebih tinggi dibandingkan penduga langsung *standardized mortality ratio*.
2. Perbaikan pendugaan resiko relatif menggunakan model Poisson-Gamma dengan peubah penyerta menghasilkan penduga Bayes empirik dengan ketelitian yang meningkat bila hubungan antara peubah penyerta dengan peubah yang menjadi perhatian dapat dimodelkan secara optimum dan berasal dari data sensus atau data administratif.

DAFTAR PUSTAKA

Agresti A. 2002. *Categorical Data Analysis*. New Jersey: John Wiley & Sons.

Bayarri MJ, Berger JO. 2004. The interplay of Bayesian and Frequentist analysis. [terhubungberkala]. <http://www.stat.duke.edu/~berger/papers/interplay.pdf> [25 Nopember 2006].

- Carlin BP, Louis TA. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman & Hall.
- Clayton D, Kaldor J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671-681.
- Dean CB, MacNab YC. 2001. Modelling of rates over a hierarchical health administrative structure. *The Canadian Journal of Statistics* 29: 405-419.
- Fay RE, Herriot RA. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74: 269-277.
- Ghosh M, Rao JNK. 1994. Small area estimation: an appraisal. *Statistical Science* 9:55-76.
- Gill J. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton: Chapman & Hall.
- Kismiantini, Notodiputro KA, Kurnia A. 2006. Risk of dengue haemorrhagic fever in Bekasi Municipality with small area approach. This paper was presented in The First International Conference on Mathematics and Statistics, Bandung, Indonesia. [June 21, 2006].
- MacNab YC *et al.* 2004. Estimation in Bayesian disease mapping. *Biometrics* 60:865-873.
- Manton *et al.* 1989. Empirical Bayes procedures for stabilizing maps of US cancer mortality rates. *Journal of the American Statistical Association* 84: 637-650.
- Marshall RJ. 1991. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics* 40:283-294.
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. London: Chapman & Hall.
- Pascutto C *et al.* 2000. Statistical issues in the analysis of disease mapping data. *Statistics in Medicine* 19:2493-2519.
- Piegorsch WW. 1990. Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics* 46:863-867.
- Pringle DG. 1995. Disease mapping: A comparative analysis of maximum likelihood and empirical Bayes estimates of disease risk. [terhubung berkala]. <http://www.nuim.ie/staff/dpringle/ebe.pdf> [27 Juli 2006].
- Rao JNK. 1999. Some recent advances in model-based small area estimation. *Survey Methodology* 25:175-186.

- Rao JNK. 2003. *Small Area Estimation*. New York: John Wiley and Sons.
- Rao JNK. 2005. Small area estimation: methodological issues and applications. Presented at the workshop "Small Area Estimation and the Local Territory", Catholic University of the Sacred Heart, Piacenza, Italy. [May 13, 2005].
- SAS Institute Inc. 2004. SAS/STAT[®] 9.1 User's Guide. [terhubung berkala]. <http://support.sas.com/> [27 Maret 2007].
- Stern HS, Cressie N. 2000. Posterior predictive model checks for disease mapping models. *Statistics in Medicine* 19:2377-2397.
- Tsutakawa RK. 1988. Mixed model for analyzing geographic variability in mortality rates. *Journal of the American Statistical Association* 83: 37-42.
- Wakefield J, Elliott P. 1999. Issues in the statistical analysis of small area health data. *Statistics in Medicine* 18:2377-2399.
- Wakefield J. 2006. Disease mapping and spatial regression with count data. [terhubung berkala]. <http://www.bepress.com/uwbiostat/paper286.pdf> [17 Juni 2006].
- Yasui *et al.* 2000. An empirical evaluation of various priors in the empirical Bayes estimation of small area disease risks. *Statistics in Medicine* 19:2409-2420.