Original Article

นิพนธ์ต้นฉบับ

ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกในประเทศไทย โดยใช้ตัวแบบเชิงคณิตศาสตร์ Generalized Estimating Equation (GEE)

สุนีย์ สัมมาทัต
นิตยา บุญสิทธิ์
กฤษฎา เหล็กดี
สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคล พระนคร

บทคัดย่อ งานวิจัยนี้มีจุดประสงค์เพื่อหาปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกในประเทศไทย โดยศึกษาตัวแบบเชิงคณิตศาสตร์ Generalized Estimating Equation (GEE) ที่ตัวแปรตามมีการแจกแจงแบบ Poisson เปรียบเทียบกับแบบ Negative Binomial ข้อมูลที่ใช้ศึกษาเป็นข้อมูลทุติยภูมิระดับจังหวัด ประกอบด้วย จำนวนผู้ป่วยโรคไข้เลือดออก ปริมาณฝน อุณหภูมิเฉลี่ย พื้นที่ป่าไม้ รายได้ครัวเรือน รวมทั้งภาค และฤดูกาล ผลการวิจัยพบว่า ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial มีความเหมาะสมมากกว่าแบบ Poisson กำหนดให้ภาคกลาง และฤดูกาลช่วงเดือน กุมภาพันธ์-เมษายน เป็นกลุ่มอ้างอิง ที่ระดับนัยสำคัญ 0.05 ปัจจัยที่มีอิทธิพลต่อความเสี่ยงของการเกิดโรคไข้เลือดออกคือ ปริมาณฝน (Relative Risk (RR) = 1.0009) อุณหภูมิเฉลี่ย (RR = 1.1736) พื้นที่ป่าไม้ (RR = 1.0482) ภาคใต้ (RR = 2.7390) ภาคตะวันออก (RR = 2.0489) ภาคตะวันตก (RR = 1.4051) ฤดูกาลช่วง พฤษภาคม-กรกฎาคม (RR = 3.0526) และ สิงหาคม-ตุลาคม (RR = 3.2677)

คำสำคัญ: **โรคไข้เลือดออก, Generalized Estimating Equations (GEE), Poisson, Negative Binomial**

บทนำ

โรคไข้เลือดออกเป็นปัญหาสาธารณสุขในหลายประเทศทั่วโลก เป็นโรคติดเชื้อไวรัสที่มียุงลายเป็นแมลงนำโรค คือยุงในจีนัส *Aedes* โดยเฉพาะ *Aedes aegypti* หรือยุงลายบ้านซึ่งระบาดในเขตร้อน ผู้ป่วยจะมีอาการไข้ ปวดศีรษะ ปวดกล้ามเนื้อ ปวดข้อ และมีผื่น เป็นที่มาของชื่ออีกชื่อหนึ่งของไข้เลือดออก คือ break-bone fever การดำเนินโรคแบ่งออกเป็น 3 ระยะ คือ ระยะไข้ ระยะวิกฤต และระยะฟื้น[1] มีการศึกษาเกี่ยวกับปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกกันอย่างแพร่หลาย เช่น อดุลย์ กล้าขยัน[2] ศึกษาปัจจัยสภาพแวดล้อมกับการเกิดโรคไข้เลือดออกในจังหวัดกาฬสินธุ์ โดยใช้ระบบสารสนเทศภูมิศาสตร์ Gharbi และคณะ[3] ศึกษาอุบัติการณ์ของโรคไข้เลือดออกในหมู่เกาะ

Guadeloupe ในทะเลแคริบเบียน โดยใช้ตัวแบบ Seasonal autore-gressive integrated moving average (SARIMA)

การใช้ตัวแบบวิเคราะห์ข้อมูลไข้เลือดออกที่มีการวัดซ้ำในประเทศไทยยังไม่มีการนำเสนอมาก่อน ตัวแบบสำหรับข้อมูลที่มีการวัดซ้ำที่มีการนำไปประยุกต์ใช้กันอย่างแพร่หลายคือ ตัวแบบเชิงคณิตศาสตร์ Generalized Estimating Equation (GEE) ซึ่งนำเสนอโดย Liang และ Zeger[4] เป็นตัวแบบที่แสดงความสัมพันธ์ระหว่างตัวแปรต้นที่มีได้มากกว่า 1 ตัว กับตัวแปรตามโดยไม่มีข้อสมมติ (assumption) ว่า ข้อมูลต้องเป็นอิสระกัน ค่าของตัวแปรตามที่มีความสัมพันธ์กันนั้น เกิดขึ้นได้เมื่อมีการเก็บข้อมูลซ้ำในหน่วยตัวอย่างเดียวกัน ตัวแปรตามเป็นได้ทั้งแบบต่อเนื่องและไม่ต่อเนื่อง สำหรับโครงสร้างความสัมพันธ์มีหลายแบบ ได้แก่ independent สมมติให้ความสัมพันธ์ระหว่างข้อมูลมีค่าเป็นศูนย์ exchangeable กำหนดให้ความสัมพันธ์มีค่าคงที่ autoregressive กำหนดให้ความสัมพันธ์มีค่าลดลงเมื่อเวลาผ่านไป และ unstructured สมมติให้ความสัมพันธ์ของข้อมูลแต่ละคู่ไม่มีรูปแบบ สมการแสดงความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามของแต่ละหน่วยตัวอย่างของตัวแบบ GEE มีค่าสัมประสิทธิ์การถดถอยเท่ากัน จึงเป็นการอธิบายขนาดอิทธิพลของปัจจัยในลักษณะภาพรวมของประชากร เรียกตัวแบบนี้ว่า Population-averaged model วนิดา ลิ่มมั่น และลี่ลี อิงศรีสว่าง[5] นำตัวแบบ GEE ไปใช้ในการหาปัจจัยที่ส่งผลต่อการเกิดอุบัติเหตุจากการจราจรทางถนน รุ้งรวี อำนาจตระกูล และลี่ลี อิงศรีสว่าง[6] นำตัวแบบ GEE ไปใช้พยากรณ์ประสิทธิภาพของหอทำน้ำเย็นชนิดลมดูดแบบไหลสวนทาง และกฤษฎา เหล็กดี และลี่ลี อิงศรีสว่าง[7] นำตัวแบบ GEE ไปใช้วิเคราะห์ความเสี่ยงของการเกิดโรคมาลาเรียในประเทศไทย

สำหรับตัวแบบ GEE เมื่อตัวแปรตามมีค่าเป็นจำนวนนับจะถูกสมมติให้มีการแจกแจงแบบ Poisson แต่ในการศึกษาจำนวนมาก พบว่าข้อมูลที่ได้จากการนับมักจะมีค่าความแปรปรวนมากกว่าค่าเฉลี่ยเรียกว่าเกิดปัญหา Over dispersion วิธีการแก้ปัญหาแบบหนึ่งคือการใช้แจกแจงแบบ Negative Binomial แทน Poisson การแจกแจงแบบ Negative Binomial ค่าความแปรปรวนมีค่ามากกว่าค่าเฉลี่ย การแจกแจงแบบ Poisson เป็นกรณีเฉพาะกรณีหนึ่งของการแจกแจงแบบ Negative Binomial การพิจารณาแก้ปัญหา Over dispersion ในการวิเคราะห์ข้อมูล จะทำให้ได้ค่าประมาณที่มีประสิทธิภาพมากขึ้น

เนื่องจากยังไม่มีการนำตัวแบบ GEE มาใช้กับข้อมูลไข้เลือดออกที่มีการวัดซ้ำในประเทศไทยมาก่อน ผู้วิจัยจึงสนใจที่จะศึกษาปัจจัยเสี่ยงในการเกิดโรคไข้เลือดออกในประเทศไทยโดยใช้ตัวแบบ GEE ปัจจัยที่นำมาพิจารณาได้แก่ ปริมาณฝน อุณหภูมิเฉลี่ย พื้นที่ป่าไม้ รายได้ครัวเรือน ภาค (กลาง เหนือ ตะวันออกเฉียงเหนือ ใต้ ตะวันออก และตะวันตก) ฤดูกาล (พฤศจิกายน-มกราคม กุมภาพันธ์-เมษายน พฤษภาคม-กรกฎาคม และสิงหาคม-ตุลาคม) ตัวแบบที่นำมาศึกษา คือตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson เปรียบเทียบกับแบบ Negative Binomial ผลที่ได้จากการศึกษามีประโยชน์สำหรับการวางแผนป้องกัน และควบคุมการเกิดโรคไข้เลือดออกในประเทศไทย

วิธีการศึกษา

1. ทฤษฎีทางสถิติ

1.1 ตัวแบบ Poisson Regression[8]

ตัวแบบ Poisson Regression มีลักษณะดังนี้ กำหนดให้ Y_i , $i = 1,2,3,....,n$ เป็นตัวแปรตามที่มีค่าเป็นจำนวนนับ มีการแจกแจงแบบ Poisson มีค่าเฉลี่ยเท่ากับ μ_i และความแปรปรวนมีค่าเท่ากับค่าเฉลี่ย นั่นคือ Y_i - $Poisson(\mu_i)$ การแจกแจงความน่าจะเป็นของ Y_i คือ

$f(y_i; \mu_i) =$, $y_i = 0,1,2,...$ และ $E(Y)_i = Var(Y_i) = \mu_i$

ให้ตัว $\boldsymbol{X_i} = (X_{i0}, X_{i1},...,X_{ip})^T$ เป็นแปรต้นที่มีความ

สัมพันธ์เชิงเส้นกับ Y_i รูปแบบความสัมพันธ์ที่ใช้กันโดยทั่วไปคือ Canonical link ที่มีรูปแบบเป็น Natural log function

กำหนดให้ $\boldsymbol{\beta}_i$, $(b_0,b_1,...,b_p)^T$ เป็นพารามิเตอร์ ความสัมพันธ์ที่ได้คือ $\log(\mu_i) = \mathbf{X}_i^T\boldsymbol{\beta}$

ในการประมาณค่า $\boldsymbol{\beta}$ นั้น สามารถใช้วิธีการของ Maximum likelihood และแก้สมการหาคำตอบโดยใช้วิธีการวนซ้ำเชิงตัวเลข (Numerical iterative method)

1.2 ตัวแบบ Negative Binomial Regression[8]

ตัวแบบ Negative Binomial Regression มีลักษณะดังนี้

กำหนดให้ มีการแจกแจงแบบ Negative Binomial โดยที่ และ, คือ Dispersion factor

$$f(y_i;\mu_i,\alpha)=\frac{\Gamma(y_i+1/\alpha)}{y_i!\Gamma(1/\alpha)}\left(1+\alpha\mu_i\right)^{-1/\alpha}\left(1+\frac{1}{\alpha\mu_i}\right)^{-y_i},\ y_i\geq 0,\ y_i=0,1,2,...$$

$\boldsymbol{\Gamma}$ คือ Gamma function และเมื่อ $\alpha\rightarrow 0$ Negative Binomial จะลู่เข้าสู่ Poisson และ $V(Y_i)\rightarrow\mu_i$

ตัวแบบ Negative Binomial Regression สามารถเขียนอยู่ในรูป $\log(\mu_i) = \mathbf{X}_i^T\boldsymbol{\beta}$ การประมาณค่า $\boldsymbol{\beta}$ สามารถใช้วิธีการของ Maximum likelihood และแก้สมการหาคำตอบโดยใช้วิธีการวนซ้ำเชิงตัวเลข เช่นเดียวกันกับตัวแบบ Poisson Regression

1.3 ตัวแบบ Generalized Estimating Equation (GEE)

GEE เป็นตัวแบบที่ใช้สำหรับตัวแปรตามที่ค่าสังเกตมีความสัมพันธ์กัน มีลักษณะดังนี้

กำหนดให้ Y_{ij}, j = 1,...,n_i i = 1,..., M แทนค่าสังเกตของหน่วยศึกษาที่ i ในครั้งที่ j

ให้ N แทนจำนวนค่าสังเกตทั้งหมด N =

Y_{ij} มีความสัมพันธ์กับตัวแปรต้น $\mathbf{X}_{ij}$ = $(X_{ij.0}, X_{ij.1},...,X_{ij.p})^T$ และพารามิเตอร์

$\boldsymbol{\beta}_i$ =$(b_0, b_1,...,b_p)^T$ ตามรูปแบบของ Link function ต่อไปนี้

Identity link: $g(\alpha) = \alpha$

Natural log link: $g(\alpha) = \log(\alpha)$

Logit link: $g(\alpha) = \log(\alpha/(1 - \alpha))$

ความสัมพันธ์ของ Y_{ij} กับ $\mathbf{X}_{ij}$ = $(X_{ij.0}, X_{ij.1},...,X_{ij.p})^T$ เขียนได้ดังนี้

$g(E(Y_{ij} \mid \mathbf{X}_{ij})) = \mathbf{X}_{ij}^T\boldsymbol{\beta}$ และเรียกความสัมพันธ์นี้ว่า Marginal regression model

Covariance Matrix ของ $\mathbf{Y}_i = (Y_{i1},...,Y_{ij})^T$ คือ $\mathbf{V}_i = \phi A_i^{1/2}\mathbf{R}(p)\mathbf{A}_i^{1/2}$

เมื่อ ø คือ Dispersion parameter $\mathbf{A}_i$ คือ Diagonal matrix of a variance function และ $\mathbf{R}(p)$ คือ Correlation matrix แสดง รูปแบบความสัมพันธ์ของข้อมูลของ $\mathbf{Y}_i = (Y_{i1},...,Y_{ij})^T$ ซึ่งมีหลายรูปแบบต่อไปนี้

Exchangeable มีโครงสร้างความสัมพันธ์คือ

$$\mathbf{R}(\rho)=\begin{bmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{bmatrix}$$

First-order autoregressive model มีโครงสร้างดังนี้

$$\mathbf{R}(\rho)=\begin{bmatrix}1 & \rho & \cdots & \rho^{j-1}\\ \rho & 1 & \cdots & \rho^{j-2}\\ \vdots & \vdots & \ddots & \vdots\\ \rho^{j-1} & \rho^{j-2} & \cdots & 1\end{bmatrix}$$

Unstructured correlation มีโครงสร้างดังนี้

$$\mathbf{R}(\rho)=\begin{bmatrix}1 & \rho_{12} & \cdots & \rho_{1j}\\ \rho_{21} & 1 & \cdots & \rho_{2j}\\ \vdots & \vdots & \ddots & \vdots\\ \rho_{j1} & \rho_{j2} & \cdots & 1\end{bmatrix}$$

การประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}$ ใช้วิธีการวนซ้ำเชิงตัวเลขในการหาคำตอบของ Score function ของ $\boldsymbol{\beta}$ ซึ่งแทนด้วย $U(\boldsymbol{\beta}) = \boldsymbol{\Sigma}\mathbf{D}_i^T\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$ โดยที่ $\mathbf{D}_i$ = , k = 1, 2,..., p มีขั้นตอนดังนี้

1) กำหนดรูปแบบของ $\mathbf{R}(p)$ คำนวณค่า $\mathbf{V}_i$ และ

b ซึ่งเป็นค่าประมาณของ **β**

2) คำนวณค่าคลาดเคลื่อน $r_{ij} = Y_{ij} - \mu_{ij}$

3) คำนวณค่า $\mathbf{V}_i$ ขึ้นมาใหม่จากค่าความคลาดเคลื่อน $\boldsymbol{r}_{ij}$ ในข้อ 2)

4) คำนวณค่า **b** ขึ้นมาใหม่จากค่าจาก $\mathbf{V}_i$ ที่ได้จากข้อ 3)

5) ทำซ้ำข้อ 2) - 4) จนกระทั่ง **b** ลู่เข้า (Convergence)

2. ระเบียบวิธีวิจัย

ประชากร คือ ผู้ป่วยโรคไข้เลือดออกของประเทศไทยในแต่ละปี

ตัวอย่าง คือ ผู้ป่วยโรคไข้เลือดออกของประเทศไทยปี 2553 จำนวน 53,149 คน

ตัวแปรต้น คือ ปริมาณฝน อุณหภูมิเฉลี่ย พื้นที่ป่าไม้ รายได้ครัวเรือน ภาค (กลาง เหนือ ตะวันออกเฉียงเหนือ ใต้ ตะวันออก และตะวันตก) ฤดูกาล (พฤศจิกายน-มกราคม กุมภาพันธ์-เมษายน พฤษภาคม-กรกฎาคม และสิงหาคม-ตุลาคม)

ตัวแปรตาม คือ จำนวนผู้ป่วยโรคไข้เลือดออกปี 2553 รายเดือนในแต่ละจังหวัดของประเทศไทย

ขั้นตอนการทำวิจัย มีดังนี้

1) ศึกษาเรื่องโรคไข้เลือดออก เช่น สาเหตุของโรค การติดต่อ อาการ การวินิจฉัย การรักษา และการป้องกัน ตัวแบบ Poisson Regression และ Negative Binomial Regression รวมทั้งตัวแบบ GEE ซึ่งเป็นตัวแบบที่ใช้วิเคราะห์ข้อมูลที่ตัวแปรตามมีการวัดซ้ำหรือมีความสัมพันธ์กัน โดยเฉพาะกรณีที่ตัวแปรตามมีค่าเป็นจำนวนนับคือมีการแจกแจงแบบ Poisson และ Negative Binomial ศึกษางานวิจัยที่เกี่ยวข้องกับการประยุกต์ใช้ตัวแบบ GEE และงานวิจัยที่เกี่ยวข้องกับโรคไข้เลือดออก

2) เก็บข้อมูลระดับจังหวัด ปี 2553ได้แก่ จำนวนผู้ป่วยไข้เลือดออกรายเดือนของแต่ละจังหวัดในประเทศไทย จำนวนประชากรในแต่ละจังหวัด ปริมาณฝน อุณหภูมิเฉลี่ย พื้นที่ป่าไม้ รายได้ครัวเรือน ภาค (กลาง เหนือ ตะวันออกเฉียงเหนือ ใต้ ตะวันออก และภาคตะวันตก) ฤดูกาล (เดือนพฤศจิกายน-มกราคม กุมภาพันธ์-เมษายน พฤษภาคม-กรกฎาคม และสิงหาคม-ตุลาคม)

จำนวนผู้ป่วยโรคไข้เลือดออกรายเดือน ได้จากกรมควบคุมโรค กระทรวงสาธารณสุข[9] พื้นที่ป่าไม้รายจังหวัด จากกรมป่าไม้[10] จำนวนประชากรปี 2553 และรายได้ครัวเรือนรายจังหวัด จากสำนักงานสถิติแห่งชาติ[11] ปริมาณฝนรายเดือน และอุณหภูมิเฉลี่ยรายเดือน จากกรมอุตุนิยมวิทยา[12] การวิเคราะห์ข้อมูลใช้โปรแกรมวิเคราะห์ข้อมูลทางสถิติวิเคราะห์ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออก โดยใช้ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson และ Negative Binomial โดยในขั้นตอนแรกพิจารณาทีละปัจจัย เลือกปัจจัยที่มีค่า p-value น้อยกว่า 0.10 ไปวิเคราะห์ต่อในขั้นตอนที่ 2 ซึ่งพิจารณาทุกปัจจัยพร้อมกัน value คำนวณมาจากค่าสถิติ ไค-สแควร์ ที่ใช้ในการทดสอบสมมติฐานค่าสัมประสิทธิ์การถดถอยของปัจจัยว่ามีค่าเท่ากับ 0 หรือไม่ รายละเอียดตัวแบบแสดงดังต่อไปนี้

ให้ Y_{ij} แทนจำนวนผู้ป่วยไข้เลือดออกในจังหวัดที่ i เดือนที่ j เมื่อ $i = 1,..., 76$ และ $j = 1,...,12$

N_i แทนจำนวนประชากรปี 2553 ของจังหวัดที่ i

$X_{ij.1}$ แทน ปริมาณฝน (มม.) ในจังหวัดที่ i เดือนที่ j

$X_{ij.2}$ แทน อุณหภูมิเฉลี่ย (องศาเซลเซียส) ในจังหวัดที่ i เดือนที่ j

$X_{i.3}$ แทน พื้นที่ป่าไม้ (ไร่) ในจังหวัดที่ i

X_{i4} แทน รายได้ครัวเรือน (บาท/ปี) ในจังหวัดที่ i

X_{i5} แทน จังหวัดที่ i อยู่ภาคเหนือ

X_{i6} แทน จังหวัดที่ i อยู่ภาคตะวันออกเฉียงเหนือ

X_{i7} แทน แทน จังหวัดที่ i อยู่ภาคใต้

X_{i8} แทน แทน จังหวัดที่ i อยู่ภาคตะวันออก

X_{i9} แทน แทน จังหวัดที่ i อยู่ภาคตะวันตก

X_{i10} แทน เดือนพฤศจิกายน-มกราคม

X_{i11} แทน พฤษภาคม-กรกฎาคม

X_{i12} แทน และสิงหาคม-ตุลาคม

β_0 คือ Intercept และ $\beta_1,\ldots,\beta_{12}$ คือค่าสัมประสิทธิ์การถด ถอยของปริมาณฝน อุณหภูมิเฉลี่ย พื้นที่ป่าไม้ รายได้ครัวเรือน ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคใต้ ภาคตะวันออก ภาคตะวันตก ฤดูกาลช่วง เดือนพฤศจิกายน-มกราคม พฤษภาคม-กรกฎาคม และสิงหาคม-ตุลาคม ตามลำดับ

กำหนดรูปแบบความสัมพันธ์ของข้อมูลในหน่วยตัวอย่างเดียวกันที่เกิดจากการวัดซ้ำ เป็นแบบ First-order autoregressive model เนื่องจากข้อมูลผู้ป่วยโรคไข้เลือดออกรายเดือนมีลักษณะเป็นข้อมูลอนุกรมเวลา ความสัมพันธ์ของตัวแปรต้นกับตัวแปรตามอยู่ในรูป

$$\log(\mu_{ij}) = \log(N_i) + \beta_0 + \beta_1 X_{ij.1} + \beta_2 X_{ij.2} + \beta_3 X_{i.3} + \beta_4 X_{i.4} + \beta_5 X_{i.5} + \beta_6 X_{i.6} + \beta_7 X_{i.7} + \beta_8 X_{i.8} + \beta_9 X_{i.9} + \beta_{10} X_{i.10} + \beta_{11} X_{i.11} + \beta_{12} X_{i.12}$$

ปัจจัยที่มีค่า p-value น้อยกว่า 0.05 เป็นปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออก ขนาดของปัจจัยอธิบายด้วยค่า Relative Risk (RR) และในการเปรียบเทียบความเหมาะสมเมื่อกำหนดให้ Y_{ij} มีการแจกแจงแบบ Poisson กับแบบ Negative Binomial นั้น ใช้ค่า Mean deviance เป็นเกณฑ์ แบบที่ Mean deviance ที่มีค่าใกล้ 1[13] จะเป็นแบบที่มีความเหมาะสมมากกว่า

ผลการศึกษา

ข้อมูลที่เก็บรวบรวมมีลักษณะดังนี้ จำนวนผู้ป่วยโรคไข้เลือดออกในประเทศไทย ปี 2553 จำนวน 53,149 คน จำนวนผู้ป่วยโรคไข้เลือดออกเฉลี่ยต่อเดือนต่อจังหวัด 58.277 คน ปริมาณฝนเฉลี่ยต่อเดือนต่อจังหวัด 133.958 มม. และอุณหภูมิเฉลี่ยต่อเดือนต่อจังหวัด 27.285 องศาเซลเซียส จำนวนประชากรเฉลี่ยต่อจังหวัด 844,443.039 คน พื้นที่ป่าไม้เฉลี่ยต่อจังหวัด 2113.345 ไร่ และ รายได้ครัวเรือนเฉลี่ยต่อปีต่อจังหวัด 114,943.842 บาท

ในการวิเคราะห์ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออก โดยใช้ตัวแบบ GEE มี 2 ขั้นตอน ขั้นตอนแรกพิจารณาทีละปัจจัย ปัจจัยที่มีค่า p-value น้อยกว่า

ตารางที่ 1 ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออก ที่วิเคราะห์ทุกปัจจัยพร้อมกัน ตัวแปรตามมีการแจกแจงแบบ Poisson

ปัจจัย	ค่าประมาณ	ส่วนเบี่ยงเบนมาตรฐาน	ไค-สแควร์	P-value (p)	ความเสี่ยงสัมพัทธ์ (RR)
Intercept	-4.225	0.8904	22.51	<0.0001	0.01463
ปริมาณฝน	0.0008	0.0002	16.46*	<0.0001	1.0008
อุณหภูมิเฉลี่ย	0.1583	0.0297	28.31*	<0.0001	1.1715
พื้นที่ป่าไม้	0.0857	0.0101	71.8*	<0.0001	1.0895
เหนือ	0.5183	0.1326	15.28*	<0.0001	1.6792
ตะวันออกเฉียงเหนือ	0.3341	0.102	10.74*	0.001	1.3967
ใต้	1.3019	0.1051	153.33*	<0.0001	3.6763
ตะวันออก	0.8718	0.1219	51.13*	<0.0001	2.3912
ตะวันตก	-0.039	0.1616	0.06	0.809	0.9618
พ.ย.-ม.ค.	0.3223	0.1568	4.23*	0.0398	1.3803
พ.ค.-ก.ค.	1.2095	0.1016	141.85*	<0.0001	3.3518
ส.ค.-ต.ค.	1.3817	0.1122	151.67*	<0.0001	3.9817

**มีนัยสำคัญทางสถิติที่ระดับ 0.05*

0.10 จะถูกนำไปวิเคราะห์ต่อในขั้นตอนที่ 2 คือวิเคราะห์ทุกปัจจัยพร้อมกัน ตัวแบบที่ใช้คือตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson เปรียบเทียบกับแบบ Negative Binomial ผลการวิเคราะห์ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกทุกปัจจัยพร้อมกัน โดยใช้ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson ผลการวิเคราะห์แสดงในตารางที่ 1

จากตารางที่ 1 ที่ระดับนัยสำคัญ 0.05 ปัจจัยที่มีอิทธิพลต่อความเสี่ยงของการเกิดโรคไข้เลือดออกคือ ที่มีค่า P-value ซึ่งคำนวณมาจากค่าสถิติ ไค-สแควร์ ที่ใช้ในการทดสอบสมมติฐานค่าสัมประสิทธิ์การถดถอยของปัจจัย ว่ามีค่าเท่ากับ 0 หรือไม่ ได้แก่ ปริมาณฝน (RR 1.0008) อุณหภูมิเฉลี่ย (RR 1.1715) พื้นที่ป่าไม้ (RR 1.0895) ภาคเหนือ (RR 1.6792) ภาคตะวันออกเฉียงเหนือ (RR 1.3967) ภาคใต้ (RR 3.6763) ภาคตะวันออก (RR 2.3912) ฤดูกาลช่วง เดือนพฤศจิกายน-มกราคม (RR 1.3803) พฤษภาคม-กรกฎาคม (RR 3.3518) และสิงหาคม-ตุลาคม (RR 3.9817) ยกเว้นภาคตะวันตก (p 0.809) ไม่มีอิทธิพลต่อความเสี่ยงของการเกิดโรคไข้เลือดออก ถ้าปริมาณฝนเพิ่มขึ้น 10 มม. ความเสี่ยงของการเป็นไข้เลือดออกจะเพิ่มขึ้น ร้อยละ 0.8 ถ้าอุณหภูมิสูงขึ้น 1 องศาเซลเซียส ความเสี่ยงจะเพิ่มขึ้น ร้อยละ 17.15 ถ้าพื้นที่ป่าไม้เพิ่มขึ้น 1 ไร่ ความเสี่ยงจะเพิ่มขึ้น ร้อยละ 8.95 กำหนดให้ภาคกลางเป็นกลุ่มอ้างอิง ภาคเหนือมีความเสี่ยงเป็น 1.6792 เท่า ภาคตะวันออกเฉียงเหนือมีความเสี่ยงเป็น 1.3967 เท่า ภาคใต้มีความเสี่ยงเป็น 3.6763 เท่า ภาคตะวันออกมีความเสี่ยงเป็น 2.3912 เท่า กำหนดให้ฤดูกาลช่วงเดือนกุมภาพันธ์-เมษายน เป็นกลุ่มอ้างอิง ฤดูกาลช่วงเดือนพฤศจิกายน-มกราคม มีความเสี่ยงเป็น 1.3803 เท่า พฤษภาคม-กรกฎาคม มีความเสี่ยงเป็น 3.3518 เท่า และช่วงเดือน สิงหาคม- ตุลาคม มีความเสี่ยงเป็น 3.9817 เท่า

ผลการวิเคราะห์ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกทุกปัจจัยพร้อมกัน โดยใช้ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial ผล

ตารางที่ 2 ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกที่วิเคราะห์ทุกปัจจัยพร้อมกัน ตัวแปรตามมีการแจกแจงแบบ Negative Binomial

ปัจจัย	ค่าประมาณ	ส่วนเบี่ยงเบนมาตรฐาน	ไค-สแควร์	P-value (p)	ความเสี่ยงสัมพัทธ์ (RR)
Intercept	-3.9386	0.7871	25.04	<.0001	0.0195
ปริมาณฝน	0.0009	0.0003	12.50*	0.0004	1.0009
อุณหภูมิเฉลี่ย	0.1601	0.0269	35.48*	<0.0001	1.1736
พื้นที่ป่าไม้	0.0471	0.0139	11.44*	0.0007	1.0482
เหนือ	0.2753	0.1409	3.82	0.0507	1.3169
ตะวันออกเฉียงเหนือ	0.0909	0.1067	0.73	0.3944	1.0951
ใต้	1.0076	0.1125	80.19*	<0.0001	2.7390
ตะวันออก	0.7173	0.1289	30.98*	<0.0001	2.0489
ตะวันตก	0.3401	0.138	6.07*	0.0137	1.4051
พ.ย.-ม.ค.	0.2211	0.1188	3.46	0.0628	1.2474
พ.ค.-ก.ค.	1.116	0.0983	128.79*	<0.0001	3.0526
ส.ค.-ต.ค.	1.1841	0.107	122.55*	<0.0001	3.2677

**มีนัยสำคัญทางสถิติที่ระดับ 0.05*

ตารางที่ 3 ผลการประเมินตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson และ Negative Binomial

Criteria For Assessing Goodness Of Fit						
	Poisson			Negative Binomial		
Criterion	DF	Value	Value/DF	DF	Value	Value/DF
Deviance	900	36633.13	40.7035	900	1043.5794	1.1595

DF = Degree of Freedom

การวิเคราะห์แสดงในตารางที่ 2

จากตารางที่ 2 ที่ระดับนัยสำคัญ 0.05 ปัจจัยที่มีอิทธิพลต่อความเสี่ยงของการเกิดโรคไข้เลือดออกคือ ปัจจัยมีค่า P-value ซึ่งคำนวณมาจากค่าสถิติ ไค-สแควร์ ที่ใช้ในการทดสอบสมมติฐานค่าสัมประสิทธิ์การถดถอยของปัจจัย ว่ามีค่าเท่ากับ 0 หรือไม่ ได้แก่ ปริมาณฝน (RR 1.0009) อุณหภูมิเฉลี่ย (RR 1.1736) พื้นที่ป่าไม้ (RR 1.0482) ภาคใต้ (RR 2.739)ภาคตะวันออก (RR 2.0489) ภาคตะวันตก (RR 1.4051) ฤดูกาลช่วงพฤษภาคม-กรกฎาคม (RR 3.0526) และ สิงหาคม-ตุลาคม (RR 3.2677) ยกเว้นภาคเหนือ (p 0.0507) ภาคตะวันออกเฉียงเหนือ (p 0.3944) และฤดูกาลช่วงเดือนพฤศจิกายน-มกราคม (p 0.0628) ไม่มีอิทธิพลต่อความเสี่ยงของการเกิดโรคไข้เลือดออก ถ้าปริมาณฝนเพิ่มขึ้น 10 มม. ความเสี่ยงของการเป็นไข้เลือดออกจะเพิ่มขึ้น ร้อยละ 0.9 ถ้าอุณหภูมิสูงขึ้น 1 องศาเซลเซียส ความเสี่ยงจะเพิ่มขึ้น ร้อยละ 17.36 ถ้าพื้นที่ป่าไม้เพิ่มขึ้น 1 ไร่ ความเสี่ยงจะเพิ่มขึ้น ร้อยละ 4.82 กำหนดให้ภาคกลางเป็นกลุ่มอ้างอิง ภาคใต้มีความเสี่ยงเป็น 2.739 เท่า ภาคตะวันออกมีความเสี่ยงเป็น 2.0489 เท่า ภาคตะวันตกมีความเสี่ยงเป็น 1.4051 เท่า กำหนดให้ฤดูกาลช่วงเดือนกุมภาพันธ์-เมษายน เป็นกลุ่มอ้างอิง ฤดูกาลช่วงเดือน พฤษภาคม-กรกฎาคม มีความเสี่ยงเป็น 3.0526 เท่า และช่วงเดือน สิงหาคม-ตุลาคม มีความเสี่ยงเป็น 3.2677 เท่า

ผลการประเมินตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson และ Negative Binomial แสดงในตารางที่ 3

จากตารางที่ 3 เมื่อใช้เกณฑ์ Mean Deviance มีค่าใกล้ 1 เป็นตัวแบบที่เหมาะสม ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial มีค่า Mean Deviance เท่ากับ 1.1595 จึงมีความเหมาะสมมากกว่าแบบ Poisson ซึ่งมีค่า Mean Deviance เท่ากับ 40.7035 ดังนั้นการวิเคราะห์ข้อมูลในงานวิจัยนี้จึงเลือกใช้ผลการวิเคราะห์ที่ได้จากตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial

วิจารณ์

ตัวแบบ GEE เป็นตัวแบบที่ใช้วิเคราะห์ข้อมูลที่ตัวแปรตามมีความสัมพันธ์กัน ซึ่งเกิดขึ้นเมื่อตัวแปรนั้นมีการวัดซ้ำ และตัวแปรตามมีความสัมพันธ์กับตัวแปรต้นในลักษณะของ Generalized Linear Model ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ตัวแปรตามคือ จำนวนผู้ป่วยไข้เลือดออกรายเดือนในแต่ละจังหวัดของประเทศไทย เนื่องจากข้อมูลผู้ป่วยในจังหวัดเดียวกันที่ถูกเก็บมาจำนวน 12 ครั้ง มีความสัมพันธ์กันในลักษณะของอนุกรมเวลา ตัวแบบ GEE จึงถูกนำมาใช้ในการวิจัยในครั้งนี้ โดยกำหนดให้ความสัมพันธ์เป็นแบบ First-order Autoregressive model ข้อมูลที่เป็นจำนวนนับเมื่อกำหนดให้มีการแจกแจงแบบ Poisson มักเกิดปัญหา Over dispersion คือความแปรปรวนมีค่าสูงกว่าเฉลี่ย ซึ่งไม่เป็นไปตามสมบัติของการแจกแจงแบบ Poisson ที่ว่าค่าเฉลี่ยมีค่าเท่ากับค่าความแปรปรวน การวิเคราะห์ข้อมูลโดยไม่พิจารณา Over dispersion จะทำให้ได้ค่า

ประมาณพารามิเตอร์ที่ไม่ถูกต้อง ในการแก้ปัญหา Over dispersion วิธีที่ใช้กันอย่างกว้างขวางคือ ใช้การแจกแจงแบบ Negative Binomial แทนการแจกแจงแบบ Poisson[8] ค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบ Negative Binomial คือ $E(Y_{ij}) = \mu_{ij}$ และ $Var(Y_{ij}) = \mu_{ij}(1 + \alpha\mu_{ij})$ เมื่อ $\alpha > 0$ คือ Dispersion factor ถ้า $\alpha > 0$ แล้ว $Var(Y_{ij}) = \mu_{ij}$ แต่ถ้า $\alpha > 0$ ความแปรปรวนจะมีค่ามากกว่าค่าเฉลี่ย ดังนั้น Poisson จึงเป็นกรณีเฉพาะกรณีหนึ่งของ Negative Binomial ผลการวิเคราะห์ข้อมูลแสดงให้เห็นว่า เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงแบบ Poisson ปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออกแตกต่างจากกรณีของการแจกแจงแบบ Negative Binomial เมื่อใช้ Mean deviance เป็นเกณฑ์ สำหรับเปรียบเทียบ สรุปได้ว่า ตัวแปรตามที่มีการแจกแจงแบบ Negative Binomial เหมาะสมกว่าแบบ Poisson การศึกษาในครั้งนี้จึงเลือกใช้ Negative Binomial ซึ่งสอดคล้องกับ Molla และ Muniswamy[14] ที่พบว่า Negative Binomial เหมาะสมกว่า Poisson เมื่อข้อมูลของตัวแปรตามเป็นจำนวนนับที่เกิด Over dispersion

ผลการศึกษาโดยใช้ Negative Binomial พบว่า ปริมาณฝน อุณหภูมิ และพื้นที่ป่าไม้ เป็นปัจจัยเสี่ยงของการเกิดโรคไข้เลือดออก โดยถ้าปริมาณเพิ่มขึ้น 10 มม. ความเสี่ยงของการเป็นไข้เลือดออกจะเพิ่มขึ้น ร้อยละ 0.9 ถ้าอุณหภูมิสูงขึ้น 1 องศาเซลเซียส ความเสี่ยงของการเกิดไข้เลือดออกจะเพิ่มขึ้นร้อยละ 17.36 และถ้าพื้นที่ป่าไม้เพิ่มขึ้น 1 ไร่ ความเสี่ยงจะเพิ่มขึ้น ร้อยละ 4.82 ทั้ง 3 ปัจจัยดังกล่าว ส่งผลให้การเกิดโรคไข้เลือดออกเพิ่มมาก ผลการศึกษานี้สอดคล้องกับ Jeefoo และคณะ[15] ที่พบว่า ปริมาณฝน และอุณหภูมิที่สูงขึ้น ทำให้เกิดไข้เลือดออกมากขึ้น

ภาคใต้มีความเสี่ยงเป็น 2.739 เท่าของภาคกลาง ภาคตะวันออกมีความเสี่ยงเป็น 2.0489 เท่า ของภาคกลาง และภาคตะวันตก มีความเสี่ยงเป็น 1.4051 ของภาคกลาง ภูมิประเทศของภาคใต้ ภาคตะวันออก และภาคตะวันตก เป็นป่าเขา ส่งผลให้เกิดฝนตกชุก เกิดแหล่งน้ำ บริเวณน้ำขัง จึงทำให้การเกิดโรคไข้เลือดออกสูงกว่าภาคกลาง ซึ่งพื้นที่ส่วนใหญ่เป็นที่ราบลุ่มมีปริมาณฝนน้อยกว่า ฤดูกาลช่วงเดือน พฤษภาคม-กรกฎาคม มีความเสี่ยงเป็น 3.0526 เท่าของช่วงเดือน กุมภาพันธ์-เมษายน และช่วงเดือน สิงหาคม- ตุลาคม มีความเสี่ยงเป็น 3.2677 เท่าของ ช่วงเดือนกุมภาพันธ์-เมษายน ช่วงพฤษภาคม-ตุลาคม เป็นช่วงฤดูฝนของประเทศไทย ซึ่งยุงลายแพร่พันธุ์ได้ดี ไข้เลือดออกจึงเกิดมากในช่วงนี้

ข้อเสนอแนะ

งานวิจัยนี้ใช้ข้อมูลระดับจังหวัด ผลการวิเคราะห์ข้อมูล จึงเป็นการแสดงให้เห็นภาพโดยรวม ในการวิจัยครั้งต่อไปควรศึกษาในระดับบุคคล หรือระดับพื้นที่ที่เล็กลง โดยเลือกพื้นที่ หรือช่วงฤดูกาลที่มีความเสี่ยงสูงของการเกิดโรคไข้เลือดออก นำตัวแบบ GEE ที่ตัวแปรตามการแจกแจงแบบ Negative Binomial ไปประยุกต์ใช้กับงานวิจัยด้านสาธารณสุขเรื่องอื่น ๆ เช่น การศึกษาผู้ป่วยโรคลมชัก โดยที่ตัวแปรตามคือจำนวนครั้งของการเกิดอาการชักในตัวผู้ป่วย ใน 1 ช่วงเวลา เป็นต้น

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณ รศ. ดวงสุดา เตโชติรส อธิการบดีมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร ผศ. ดร. อมรา อมรแก้ว คณบดีคณะวิทยาศาสตร์และเทคโนโลยี และ ผศ. จุฑามาศ พีรพัชระ ผู้อำนวยการสถาบันวิจัยและพัฒนา ที่ ให้การสนับสนุนการทำงานวิจัยของอาจารย์ในมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร และ ขอขอบคุณ กรมควบคุมโรค กรมป่าไม้ และสำนักงานสถิติแห่งชาติ ที่เผยแพร่ข้อมูลที่ใช้ในการวิจัยในครั้งนี้

เอกสารอ้างอิง

1. วิกิพีเดีย. ไข้เลือดออกเด็งกี [online] 2555 [สืบค้น 20 มกราคม 2555]. Available from: URL: http://th.wikipedia.org/wiki/ไข้เลือดออก

2. อดุลย์ กล้าขยัน. การศึกษาปัจจัยสภาพแวดล้อมกับการเกิดโรคไข้เลือดออกในจังหวัดกาฬสินธุ์ โดยประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ (วิทยานิพนธ์ สาธารณสุขศาสตรมหาบัณฑิต). คณะสาธารณสุขศาสตร์. บัณฑิตวิทยาลัย. มหาสารคาม: มหาวิทยาลัยมหาสารคาม. 2549.
3. Gharbi M, Quenel P, Gustave J, Cassadou S, Ruche GL, Girdary L, et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. BMC Infectious Diseases 2011;11(166):1-13.
4. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13-22.
5. วนิดา ลิ่มมั่น, ลี่ลี อิงศรีสว่าง. การศึกษาปัจจัยที่ส่งผลต่อการเกิดอุบัติเหตุจากการจราจรทางถนนโดยใช้ตัวแบบ Generalized Estimating Equations และ Generalized Linear Mixed Models. พระจอมเกล้าพระนครเหนือ 2553; 20(2): 311-21.
6. รุ้งรวี อำนาจตระกูล, ลี่ลี อิงศรีสว่าง. การสร้างตัวแบบเพื่อพยากรณ์ประสิทธิภาพของหอทำน้ำเย็นชนิดลมดูดแบบไหลสวนทางด้วยวิธี GEE. วิทยาศาสตร์บูรพา 2555;17(1):87-96.
7. กฤษฎา เหล็กดี, ลี่ลี อิงศรีสว่าง. ปัจจัยเสี่ยงของการเกิดโรคมาลาเรียในประเทศไทย โดยใช้ตัวแบบ Generalized Estimating Equation (GEE) และ Generalized Linear Mixed Model (GLMM). วารสารวิชาการสาธารณสุข 2553;19(3): 364-73.
8. McCullagh P, Nelder J. Generalized Linear Models. 2nd ed. Boca Raton: Chapman and Hall/CRC; 1989.
9. สำนักระบาดวิทยา. โรคไข้เลือดออก สรุปรายงานการเฝ้าระวังโรค 2553 [online] 2554 [สืบค้น 20 มกราคม 2555]. แหล่งข้อมูล: URL: http://203.157.15.4/Annual/ANNUAL2550/Part1/ 0550_Dengue.
10. กรมป่าไม้. เนื้อที่ป่าไม้ของประเทศไทย [online] 2554 [สืบค้น 20 มกราคม 2555]. แหล่งข้อมูล: URL: http://www.forest.go.th/home/index.asp
11. สำนักงานสถิติแห่งชาติ. ข้อมูลสถิติ. [online] 2554 [สืบค้น 20 มกราคม 2555]. แหล่งข้อมูล: URL: http://service.nso.go.th/nso/nsopublish/service/servstat.html
12. กรมอุตุนิยมวิทยา. ปริมาณฝนและอุณหภูมิปี 2553 [online] 2554 [สืบค้น 20 มกราคม 2555]. แหล่งข้อมูล: URL: http://www.tmd.go.th/index.php
13. Schabenberger O. Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models. [online] [cited 20 February 2012]. Available from: URL: http://www2.sas.com/proceedings/sugi30/196-30.pdf
14. Molla DT, Muniswamy B. Power of tests for overdispersion parameter in negative binomial regression model. Journal of Mathematics 2012;1(4):29-36.
15. Jeefoo P, Tripathi NK, Souris M. Spatio-temporal diffusion pattern and hotspot detection of dengue in Chachoengsao province, Thailand. International Journal of Environment Research and Public Health 2011; 8(1):51-74.

Abstract **Risk Factors for Dengue Fever in Thailand Using Mathematical Generalized Estimating Equation (GEE)**

Sunee Sammatat, Nittaya Boonsith, Krisada Lekdee

Department of Mathematics and Statistics, Faculty of Science and Technology, Rajamangala University of Technology Phra Nakhon

Journal of Health Science **2013; 22:566-575.**

The objective of this research is to investigate risk factors for Dengue fever in Thailand. The mathematical generalized estimating equation (GEE) whose dependent variables have Poisson distributions and negative binomial distributions are studied. The secondary province-level data consisting of the number of Dengue fever patients, rainfall, average temperature, forest area, household income are collected. The regions and seasons are also considered. The research finds that the GEE whose dependent variables have negative binomial distributions are more appropriate, hence it is used for the data analysis. The central region and the season during February-April are assigned to be reference groups. At the significance of 0.05, the risk factors for Dengue fever are rainfall (Relative Risk (RR) 1.0009), average temperature (RR 1.1736), forest area (RR 1.0482), southern region (RR 2.7390), eastern region (RR 2.0489), western region (RR 1.4051), the season during May-July (RR 3.0526) and August-October (RR 3.2677).

Key words: **Dengue fever, Generalized Estimating Equation (GEE), Poisson, Negative binomial**