

Original Article

นิพนธ์รัตน์ฉบับ

# ปัจจัยเสี่ยงของการเกิดโรคมาลาเรียในประเทศไทย โดยใช้ตัวแบบ Generalized Estimating Equation (GEE) และ Generalized Linear Mixed Model (GLMM)

กฤษฎา เหล็กดี  
ลีลี อิงศรีสว่าง

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

## บทคัดย่อ

การวิจัยครั้งนี้มีจุดประสงค์เพื่อหาปัจจัยเสี่ยงของการเกิดโรคมาลาเรียในประเทศไทยด้วยตัวแบบทางสถิติที่เหมาะสม โดยศึกษาตัวแบบ Generalized Estimating Equation (GEE) และ Generalized Linear Mixed Model (GLMM) ที่ตัวแปรตามมีการแจกแจงแบบ Poisson เปรียบเทียบกับแบบ Negative Binomial ข้อมูลที่ใช้ศึกษาเป็นข้อมูลทุติยภูมิ ระดับจังหวัด ปี 2550 ได้จากหลายหน่วยงาน ตัวแปรตามคือจำนวนผู้ป่วยโรคมาลาเรียรายเดือน ตัวแปรต้นคือ ภาค ชายแดน อุดรกาล พื้นที่ป่าไม้ ปริมาณฝน อุณหภูมิเฉลี่ย และรายได้เฉลี่ยต่อหัว การคัดเลือกตัวแบบที่เหมาะสมพิจารณาจากค่าสถิติ mean deviance (กรณีของ GEE) และ mean generalized chi-square (กรณีของ GLMM) ผลการวิจัยพบว่า ทั้งตัวแบบ GEE และ GLMM ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial มีความเหมาะสมกว่าแบบ Poisson เมื่อพิจารณาในลักษณะภาพรวมทั้งประเทศจากตัวแบบ GEE พบว่าปัจจัยที่มีความสัมพันธ์กับอัตราการป่วยโรคมาลาเรีย อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ได้แก่ ภาค (ตะวันออก และใต้) ชายแดน (ติดกับพม่า ติดกับมาเลเซีย ติดกับกัมพูชา และติดกับลาว) อุดรกาล (พ.ค.-ก.ค. พ.ย.-ม.ค. และ ส.ค.-ต.ค.) อุณหภูมิเฉลี่ย และปริมาณฝน แต่เมื่อพิจารณาในลักษณะรายจังหวัดจากตัวแบบ GLMM พบว่าปัจจัยที่มีความสัมพันธ์กับอัตราการป่วยโรคมาลาเรีย ได้แก่ ภาค (ตะวันออก ใต้ และอีสาน) ชายแดน (ติดกับพม่า ติดกับมาเลเซีย และติดกับกัมพูชา) และอุดรกาล (พ.ค.-ก.ค. และพ.ย.-ม.ค.)

## คำสำคัญ:

มาลาเรีย, Generalized Estimating Equation (GEE), Generalized Linear Mixed Model (GLMM), Poisson, Negative Binomial

## บทนำ

องค์การอนามัยโรค (WHO)<sup>(1)</sup> รายงานว่าประมาณร้อยละ 40 ของประชากรโลกโดยเฉพาะประชากรใน

ประเทศยากจนมีความเสี่ยงต่อการเป็นโรคมาลาเรียสูง ในแต่ละปีมีประชากรโลกมากกว่า 500 ล้านคนที่ป่วยเป็นมาลาเรียขั้นรุนแรง ประชากรที่ป่วยและเสียชีวิต

ส่วนใหญ่อยู่ในทวีปแอฟริกา บริเวณกึ่งทะเลทราย สะฮารา อย่างไรก็ตามพบว่ามีกรณีติดเชื้อมาลาเรียทั้งในทวีปเอเชีย ลาตินอเมริกา ตะวันออกกลาง และบางส่วนของยุโรป สำหรับประเทศไทย มาลาเรียยังคงเป็นปัญหาสาธารณสุขที่สำคัญ สำนักโรคระบาดวิทยา<sup>(2)</sup> สรุปสถานการณ์มาลาเรียในประเทศไทยว่า ตั้งแต่ พ.ศ. 2546 - 2550 อัตราการป่วยเพิ่มขึ้นเฉลี่ยปีละ 5.91 ต่อประชากรแสนคน แหล่งของเชื้อมาลาเรียอยู่บริเวณป่าเขา ชายแดนของประเทศ โดยเฉพาะบริเวณชายแดน ไทย-พม่า และไทย-กัมพูชา พบผู้ป่วยได้ตลอดทั้งปี แต่จะพบมากในช่วงฤดูฝน

มีการศึกษาเกี่ยวกับปัจจัยเสี่ยงของการเกิดโรคมาลาเรียอย่างแพร่หลาย ตัวอย่างเช่น Yeshiwondim และคณะ<sup>(3)</sup> ศึกษาอุบัติการณ์ของโรคมาลาเรียใน 543 หมู่บ้าน ในเมือง East Shoa ประเทศเอธิโอเปีย โดยใช้ตัวแบบ Poisson regression พบว่า เพศ และอายุ มีความสัมพันธ์กับอุบัติการณ์ของโรคมาลาเรีย Tian และคณะ<sup>(4)</sup> ศึกษาผลกระทบของการเปลี่ยนแปลงของสภาพอากาศต่อการแพร่เชื้อโรคมาลาเรียในเขตพื้นที่ป่าของเมือง Mengla ซึ่งอยู่ฝั่งตะวันตกเฉียงใต้ของประเทศจีน โดยใช้ตัวแบบ ARIMA พบว่าความถี่ของวันที่มีหมอกที่บ่งชี้ความสัมพันธ์สูงกับอุบัติการณ์ของโรคมาลาเรียในพื้นที่ป่าหนาที่เขตร้อนมีฝนตกมาก รัศมีศรีจันทร์<sup>(5)</sup> ศึกษาปัจจัยเสี่ยงของการติดเชื้อมาลาเรียในพื้นที่ชายแดนติดกับพม่าที่จังหวัดระนอง โดยใช้ตัวแบบ Logistic regression พบว่า อาชีพ บุคคลในครอบครัวที่มีประวัติติดเชื้อมาลาเรีย การพักผ่อนในป่า และสถานที่ทำงานอยู่ใกล้แหล่งน้ำ มีความสัมพันธ์กับการติดเชื้อมาลาเรีย ถึงแม้ว่ามีการศึกษาเกี่ยวกับปัจจัยเสี่ยงของการเกิดโรคมาลาเรียกันอย่างกว้างขวาง แต่การใช้ตัวแบบกับข้อมูลมาลาเรียที่มีการวัดซ้ำยังอยู่ในวงจำกัด ตัวแบบทางสถิติสำหรับข้อมูลที่มีการวัดซ้ำที่มีการนำไปประยุกต์ใช้กันมากคือ ตัวแบบ Generalized Estimating Equations (GEE) และ Generalized Linear Mixed Model (GLMM)

ตัวแบบ GEE นำเสนอโดย Liang และ Zeger<sup>(6)</sup> เป็นตัวแบบที่แสดงความสัมพันธ์ระหว่างตัวแปรต้นที่มีได้มากกว่า 1 ตัวกับตัวแปรตาม โดยไม่มีข้อสมมติ (assumption) ว่า ข้อมูลของตัวแปรตามต้องเป็นอิสระกัน นั่นคือยอมให้ข้อมูลของตัวแปรตามมีความสัมพันธ์กัน ซึ่งเกิดขึ้นได้เมื่อมีการเก็บข้อมูลซ้ำในหน่วยตัวอย่างเดียวกัน ตัวแปรตามเป็นได้ทั้งแบบต่อเนื่องและไม่ต่อเนื่อง รูปแบบความสัมพันธ์ของข้อมูลในหน่วยตัวอย่างเดียวกันที่เกิดจากการวัดซ้ำมีหลายรูปแบบ ได้แก่ independent, exchangeable, autoregressive และ unstructured สำหรับโครงสร้างแบบ independent สมมติให้ความสัมพันธ์ระหว่างค่าของข้อมูลวัดซ้ำมีค่าเป็นศูนย์ exchangeable กำหนดให้ ความสัมพันธ์ระหว่างค่าวัดซ้ำมีค่าคงที่ autoregressive กำหนดให้ ความสัมพันธ์ระหว่างค่าวัดซ้ำมีค่าลดลงเมื่อเวลามากขึ้น ส่วน unstructured สมมติให้ความสัมพันธ์ของข้อมูลแต่ละคู่ไม่มีรูปแบบ<sup>(7)</sup> สมการแสดงความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามของแต่ละหน่วยตัวอย่างมีค่าสัมประสิทธิ์การถดถอยเท่ากัน เป็นการอธิบายขนาดอิทธิพลของปัจจัยในลักษณะภาพรวมของประชากร จึงจัดตัวแบบนี้อยู่ในประเภท population-averaged model สำหรับตัวแบบ Generalized Linear Mixed Model (GLMM) เป็นตัวแบบที่แสดงความสัมพันธ์ระหว่างตัวแปรต้นที่มีได้มากกว่า 1 ตัวกับตัวแปรตาม และยอมให้ข้อมูลของตัวแปรตามมีความสัมพันธ์กันได้ ใช้วิเคราะห์ข้อมูลที่มีการวัดซ้ำเช่นเดียวกันกับตัวแบบ GEE ต่างกันที่มีเทอมที่เป็นอิทธิพลสุ่ม (random effect) อยู่ในตัวแบบด้วย ทำให้ค่าสัมประสิทธิ์การถดถอยในสมการของแต่ละหน่วยตัวอย่างแตกต่างกัน เป็นการอธิบายขนาดอิทธิพลของปัจจัยในลักษณะเฉพาะตัวของแต่ละหน่วยตัวอย่าง จึงจัดตัวแบบนี้อยู่ในประเภท subject-specific model ทั้งตัวแบบ GEE และ GLMM เมื่อตัวแปรตามมีค่าเป็นจำนวนนับมักจะถูกสมมติให้มีการแจกแจงแบบ Poisson แต่ในการศึกษาจำนวนมากพบว่าข้อมูลที่ได้จากการนับมักจะมีค่าความแปรปรวน

มากกว่าค่าเฉลี่ย เรียกว่าเกิด over dispersion แก้ปัญหาได้โดยการใช้แจกแจงแบบ Negative Binomial แทนการแจกแจงแบบ Poisson จะทำให้ได้ตัวประมาณ ที่มีประสิทธิภาพมากกว่า<sup>(8)</sup> ในการวิเคราะห์ข้อมูลจึงควรพิจารณาปัญหา over dispersion ด้วย

ในประเทศไทยยังไม่มีการใช้ตัวแบบทางสถิติกับข้อมูลมาลาเรียที่มีการวัดซ้ำมาก่อน ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาปัจจัยเสี่ยงของการเกิดโรคมาลาเรียโดยใช้ตัวแบบที่เหมาะสมกับข้อมูลจำนวนผู้ป่วยโรคมาลาเรียรายเดือนของแต่ละจังหวัด ของสำนักกระบาดวิทยาปี 2550 ซึ่งเป็นข้อมูลที่มีการวัดซ้ำ ตัวแบบที่พิจารณาคือตัวแบบ GEE และ GLMM ที่ตัวแปรตามมีการแจกแจงแบบ Poisson เปรียบเทียบกับแบบ Negative Binomial ผลที่ได้มีประโยชน์สำหรับการวางแผนป้องกันและควบคุมการเกิดโรคมาลาเรียในประเทศไทย

### วิธีการศึกษา

งานวิจัยนี้เป็นการวิเคราะห์ข้อมูลทฤษฎีโดยใช้ตัวแบบ GEE และ GLMM มีวิธีการดังนี้

**1. ข้อมูลและแหล่งข้อมูล** ตัวแปรตาม คือ จำนวนผู้ป่วยโรคมาลาเรียในประเทศไทยรายเดือน ปี 2550 ตัวแปรต้น คือ ภาค ชายแดน ฤดูกาล พื้นที่ป่าไม้ ปริมาณฝน อุณหภูมิเฉลี่ย และรายได้เฉลี่ยต่อหัว จำนวนผู้ป่วยโรคมาลาเรียในประเทศไทยรายเดือน ปี 2550 และจำนวนประชากรกลางปี 2550 ได้จากสำนักกระบาดวิทยา<sup>(9)</sup> พื้นที่ป่าไม้ปี 2549 ได้จากกรมป่าไม้<sup>(10)</sup> ปริมาณฝนรายเดือนปี 2550 และอุณหภูมิเฉลี่ยรายเดือนได้จากกรมอุตุนิยมวิทยา<sup>(11)</sup> และรายได้เฉลี่ยต่อหัว ปี 2550 ได้จากสำนักงานสถิติแห่งชาติ<sup>(12)</sup>

**2. การวิเคราะห์ข้อมูล** กำหนดให้  $i$  แทนจังหวัด มีค่าตั้งแต่ 1,2,..., 76 และ  $j$  แทนเดือนมีค่าตั้งแต่ 1,2,3,..., 12

$Y_{ij}$  คือจำนวนผู้ป่วยมาลาเรียในจังหวัดที่  $i$  เดือนที่  $j$

$N_i$  แทนจำนวนประชากรกลางปี 2550 ของ

จังหวัดที่  $i$

$X_{i1}$  แทน ภาค ของจังหวัดที่  $i$  กำหนดให้ 1 = เหนือ 2 = อีสาน 3 = ใต้ 4 = ตะวันออก 5 = ตะวันตก 6 = กลาง (กลุ่มอ้างอิง)

$X_{i2}$  แทน ชายแดน ของจังหวัดที่  $i$  กำหนดให้ 1 = ติดกับลาว 2 = ติดกับกัมพูชา 3 = ติดกับมาเลเซีย 4 = ติดกับพม่า 5 = ไม่ติดชายแดน (กลุ่มอ้างอิง)

$X_{i3}$  แทน ฤดูกาล ของเดือนที่  $j$  กำหนดให้ 1 = พ.ย.-ม.ค. 2 = ส.ค.-ต.ค. 3 = พ.ค.-ก.ค. 4 = ก.พ.-เม.ย. (กลุ่มอ้างอิง)

$X_{i4}$  แทน พื้นที่ป่าไม้ (ตร.กม.) ของจังหวัดที่  $i$

$X_{i5}$  แทน ปริมาณฝน (มม.) ของจังหวัดที่  $i$  เดือนที่  $j$

$X_{i6}$  แทน อุณหภูมิเฉลี่ย (Celsius) ของจังหวัดที่  $i$  เดือนที่  $j$

$X_{i7}$  แทน รายได้เฉลี่ยต่อหัว (บาท/ปี) ของจังหวัดที่  $i$

$\beta_0$  คือค่า intercept  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$  คือค่าสัมประสิทธิ์การถดถอยที่แสดงขนาดอิทธิพลของปัจจัยภาค ชายแดน ฤดูกาล พื้นที่ป่าไม้ ปริมาณฝน อุณหภูมิเฉลี่ย และรายได้เฉลี่ยต่อหัว ตามลำดับ

กำหนดรูปแบบความสัมพันธ์ของข้อมูลในหน่วยตัวอย่างเดียวกันที่เกิดจากการวัดซ้ำของทุกตัวแบบที่ศึกษาเป็นแบบ Autoregressive (AR(1)) เนื่องจากข้อมูลผู้ป่วยโรคมาลาเรียรายเดือนมีลักษณะเป็นอนุกรมเวลา<sup>(13-15)</sup>

### 2.1 ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson

ให้  $Y_{ij}$  มีการแจกแจงแบบ Poisson โดยที่  $E(Y_{ij}) = V(Y_{ij}) = \mu_{ij}$

$$Y_{ij} = \frac{\exp(\mu_{ij}) \mu_{ij}^{Y_{ij}}}{Y_{ij}!} + \varepsilon_{ij}, Y_{ij} = 0, 1, 2, 3, \dots$$

ให้  $N_i$  เป็น exposure (ตัวแปร Offset) และ  $\lambda_{ij} = \frac{\mu_{ij}}{N_i}$  เรียก  $\lambda_{ij}$  ว่า อัตรา (rate)

กำหนด Link = ln, Scale = deviance ตัวแบบ

GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson คือ

$$\ln(\lambda_{ij}) = \ln\left(\frac{\mu_{ij}}{N_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7}$$

## 2.2 ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial<sup>(3)</sup>

ให้  $Y_{ij}$  มีการแจกแจงแบบ Negative Binomial โดยที่  $E(Y) = \mu_{ij}$  และ  $V(Y) = \mu_{ij} + \frac{\mu_{ij}^2}{r_{ij}}$  เรียก  $r_{ij}$  ว่า dispersion parameter

$$Y_{ij} = \left\{ \frac{r_{ij}}{r_{ij} + \mu_{ij}} \right\}^{r_{ij}} \frac{\Gamma(r_{ij} + Y_{ij})}{\Gamma(Y_{ij} + 1) \Gamma(r_{ij})} \left\{ \frac{\mu_{ij}}{r_{ij} + \mu_{ij}} \right\}^{Y_{ij}} + \varepsilon_{ij}, Y_{ij} = 0, 1, 2, 3, \dots$$

$\Gamma$  = gamma function เมื่อ  $r_{ij}$  มีขนาดใหญ่ Negative Binomial จะลู่เข้าสู่ Poisson ( $V(Y_{ij}) \rightarrow \mu_{ij}$ )

ให้  $N_i$  เป็น exposure (ตัวแปร offset) และ  $\lambda_{ij} = \frac{\mu_{ij}}{N_i}$  เรียก  $\lambda_{ij}$  ว่า อัตรา (rate)

กำหนด Link = ln ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial คือ

$$\ln(\lambda_{ij}) = \ln\left(\frac{\mu_{ij}}{N_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7}$$

## 2.3 ตัวแบบ GLMM ที่มี random intercept และตัวแปรตามมีการแจกแจงแบบ Poisson และ Negative Binomial

ตัวแบบ GLMM ที่มี random intercept และตัวแปรตามมีการแจกแจงแบบ Poisson และ Negative Binomial มีลักษณะเหมือนตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson และ Negative Binomial ตามลำดับ เพียงแต่เพิ่มเทอมที่เป็น random intercept ( $V$ ) ที่มีการแจกแจงแบบปกติ ค่าเฉลี่ยเป็น 0 และความแปรปรวน  $\sigma_v^2$  เข้าไว้ในตัวแบบด้วย ตัวแบบ GLMM ที่มี random intercept มีลักษณะดังนี้

$$\ln(\lambda_{ij}) = \ln\left(\frac{\mu_{ij}}{N_i}\right) = (\beta_0 + V_i) + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7}$$

## 2.4 การประมาณค่าพารามิเตอร์ในตัวแบบ GEE และตัวแบบ GLMM

การประมาณค่าพารามิเตอร์ใช้โปรแกรม SAS ของภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์ ตัวแบบ

GEE ใช้ Proc Genmod ตัวแบบ GLMM ใช้ Proc Glimmix การวิเคราะห์ข้อมูลเริ่มจากการใช้ตัวแบบ GEE และ GLMM แบบ univariate ตัดตัวแปรที่มีค่า p-value มากกว่า 0.10 ออก<sup>(16)</sup> นำตัวแปรที่เหลือไปวิเคราะห์แบบ multivariate เปรียบเทียบความเหมาะสมของตัวแบบที่ตัวแปรตามมีการแจกแจงแบบ Poisson กับแบบ Negative Binomial โดยพิจารณาจากค่า mean deviance ที่มีค่าใกล้ 1<sup>(17)</sup> (กรณี GEE) และจากค่า mean generalize chi-square ที่มีค่าใกล้ 1<sup>(18)</sup> (กรณี GLMM) แล้วคัดเลือกตัวแปรโดยวิธี backward elimination ซึ่งเป็นเทคนิคที่ใช้การตัดตัวแปรออกทีละตัวหลังจากที่กำหนดเริ่มต้นให้ตัวแบบมีตัวแปรต้นครบทุกตัว

## ผลการศึกษา

**1. ตัวแบบ GEE** ผลการวิเคราะห์ข้อมูลด้วยตัวแบบ GEE แบบ univariate พบว่า ตัวแปรรายได้เฉลี่ยต่อหัว มีค่า p-value มากกว่า 0.10 จึงถูกตัดออก และเมื่อนำตัวแปรที่เหลือไปวิเคราะห์แบบ multivariate พบว่า mean deviance ของตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Poisson มีค่าสูงกว่าแบบ Negative Binomial (Poisson: mean deviance = 42.850; negative binomial: mean deviance = 1.068) เมื่อใช้เกณฑ์ Mean Deviance มีค่าใกล้ 1<sup>(17)</sup> จะได้ Negative Binomial มีความเหมาะสมกว่า Poisson ดังนั้นจึงเลือกใช้ตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial แล้วคัดเลือกตัวแปรต่อด้วยวิธี backward elimination ได้ตัวแบบที่มีค่า mean deviance เท่ากับ 1.066 ค่าประมาณพารามิเตอร์แสดงในตารางที่ 1

จากตารางที่ 1 พบว่า ปัจจัยที่มีความสัมพันธ์กับอัตราการป่วยโรคมะเร็ง อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ได้แก่ ภาค (ตะวันออก และใต้) ชายแดน (ติดกับพม่า ติดกับมาเลเซีย ติดกับกัมพูชา และติดกับลาว) ฤดูกาล (พ.ค.-ก.ค. พ.ย.-ม.ค. และส.ค.-ต.ค.) ปริมาณฝน และอุณหภูมิเฉลี่ย เมื่อเปรียบเทียบความเสี่ยงของการ

ตารางที่ 1 ค่าประมาณพารามิเตอร์ของตัวแบบ GEE ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial

พารามิเตอร์	ค่าประมาณ	ค่าคลาดเคลื่อน มาตรฐาน	ช่วงความ เชื่อมั่น 95%		ค่าสถิติ Z	Pr >  Z	Relative Risk (RR)
intercept	-15.98	0.68	-17.31	-14.66	-23.64*	<.0001	-
<b>ภาค</b>							
เหนือ	0.44	0.49	-0.52	1.39	0.90	0.37	1.55
อีสาน	-0.73	0.43	-1.57	0.11	-1.70	0.09	0.48
ใต้	2.65	0.59	1.50	3.79	4.52*	<.0001	14.10
ตะวันออก	2.65	0.60	1.48	3.82	4.43*	<.0001	14.18
ตะวันตก	-0.16	0.58	-1.30	0.98	-0.28	0.78	0.85
กลาง	0.00	0.00	0.00	0.00	-	-	1
<b>ชายแดนติดต่อ</b>							
ลาว	1.17	0.48	0.23	2.11	2.45*	0.01	3.22
กัมพูชา	1.74	0.50	0.75	2.73	3.46*	0.0005	5.71
มาเลเซีย	2.28	0.74	0.83	3.74	3.08*	0.002	9.80
พม่า	4.08	0.54	3.02	5.13	7.59*	<.0001	58.87
ไม่ติด	0.00	0.00	0.00	0.00	-	-	1
<b>ฤดูกาล</b>							
พ.ย.-ม.ค.	0.48	0.13	0.23	0.74	3.72*	0.0002	1.62
ส.ค.-ต.ค.	0.37	0.16	0.05	0.68	2.26*	0.02	1.44
พ.ค.-ก.ค.	0.70	0.21	0.28	1.11	3.28*	0.001	2.01
ก.พ.-เม.ย.	0.00	0.00	0.00	0.00	-	-	1
ปริมาณฝน	0.001	0.00	0.00	0.00	2.65*	0.008	1.0006
อุณหภูมิเฉลี่ย	0.09	0.02	0.05	0.14	3.89*	<.0001	1.10

\*มีนัยสำคัญทางสถิติที่ระดับ 0.05

หมายเหตุ: Relative Risk (RR) คือความเสี่ยงหรือโอกาสของการป่วยโรคมalariaเปรียบเทียบกับกลุ่มอ้างอิง

ป่วยโรคมalariaเรียกกับผู้ที่อยู่ภาคกลาง พบว่าความเสี่ยงของผู้ที่อยู่ในภาคตะวันออกสูงเป็น 14.18 เท่า และภาคใต้สูงเป็น 14.10 เท่า เมื่อเปรียบเทียบกับผู้ที่อยู่ในจังหวัดไม่ติดชายแดน พบว่าความเสี่ยงของผู้ที่อยู่ในจังหวัดชายแดนติดกับพม่าสูงเป็น 58.87 เท่า ชายแดนติดกับมาเลเซียสูงเป็น 9.80 เท่า ชายแดนติดกับกัมพูชาสูงเป็น 5.71 เท่า และชายแดนติดกับลาวสูงเป็น 3.22 เท่า เมื่อเปรียบเทียบกับฤดูกาลช่วงเดือน ก.พ.-เม.ย

พบว่าความเสี่ยงในฤดูกาลช่วงเดือน พ.ค.-ก.ค. สูงเป็น 2.01 เท่า ช่วงเดือน พ.ย.-ม.ค. สูงเป็น 1.62 เท่า และช่วงเดือน ส.ค.-ต.ค. สูงเป็น 1.44 เท่า เมื่ออุณหภูมิเฉลี่ยสูงขึ้น 1 หน่วย ความเสี่ยงจะเพิ่มขึ้น 1.0006 เท่า และเมื่อปริมาณฝนเพิ่มขึ้น 1 หน่วย ความเสี่ยงจะเพิ่มขึ้น 1.10 เท่า

**2. ตัวแบบ GLMM** ผลการวิเคราะห์ข้อมูลด้วยตัวแบบ GLMM แบบ univariate พบว่า ตัวแปรต้นทุก

ตัวมีค่า p-value น้อยกว่า 0.10 จึงถูกนำไปวิเคราะห์ต่อในแบบ multivariate ผลการวิเคราะห์พบว่า mean generalized chi-square ของตัวแบบ GLMM ที่ตัวแปรตามมีการแจกแจงแบบ Poisson มีค่าสูงกว่าแบบ Negative Binomial (Poisson: mean generalized chi-square = 11.96; Negative Binomial: mean generalized chi-square = 0.99) เมื่อใช้เกณฑ์ mean generalized chi-square มีค่าใกล้ 1<sup>(18)</sup> จะได้ Negative Binomial มีความเหมาะสมกว่า Poisson

ดังนั้นจึงเลือกใช้ตัวแบบ GLMM ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial แล้วคัดเลือก

ตัวแปรด้วยวิธี backward elimination ได้ตัวแบบที่มีค่า mean generalized chi-square เท่ากับ 0.99 ค่าประมาณพารามิเตอร์แสดงในตารางที่ 2

จากตารางที่ 2 พบว่า ปัจจัยที่มีความสัมพันธ์กับอัตราการป่วยโรคมาลาเรีย อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ได้แก่ ภาค (ตะวันออกเฉียงใต้ และอีสาน) ชายแดน (ติดกับพม่า ติดกับมาเลเซีย และติดกับกัมพูชา) ฤดูกาล (พ.ค.-ก.ค. และพ.ย.-ม.ค.) เมื่อเปรียบเทียบความเสี่ยงของการป่วยโรคมาลาเรียกับผู้ที่อยู่ภาคกลาง พบว่าความเสี่ยงของผู้ที่อยู่ภาคตะวันออกเฉียงใต้สูงเป็น 10.69 เท่า ภาคใต้สูงเป็น 8.78 เท่า และภาคอีสานสูงเป็น 0.33

ตารางที่ 2 ค่าประมาณพารามิเตอร์ในตัวแบบ GLMM ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial

พารามิเตอร์	ค่าประมาณ	ค่าคลาดเคลื่อนมาตรฐาน	ค่าสถิติ t	Pr >  t	Relative Risk (RR)
intercept	-13.44	0.37	-36.38*	<.0001	-
<b>ภาค</b>					
เหนือ	0.05	0.54	0.10	0.92	1.06
อีสาน	-1.11	0.53	-2.09*	0.04	0.33
ใต้	2.17	0.57	3.83*	0.0003	8.78
ตะวันออกเฉียง	2.37	0.60	3.92*	0.0002	10.69
ตะวันตก	0.16	0.76	0.21	0.83	1.18
กลาง	0.00	.	.	.	1
<b>ชายแดนติดต่อ</b>					
ลาว	0.79	0.50	1.60	0.11	2.21
กัมพูชา	1.52	0.74	2.05*	0.04	4.56
มาเลเซีย	2.28	0.79	2.89*	0.01	9.73
พม่า	3.85	0.58	6.68*	<.0001	46.77
ไม่ติด	0.00	.	.	.	1
<b>ฤดูกาล</b>					
พ.ย.-ม.ค.	-0.28	0.09	-2.99*	0.003	0.76
ส.ค.-ต.ค.	-0.09	0.09	-0.95	0.34	0.92
พ.ค.-ก.ค.	0.76	0.09	8.77*	<.0001	2.14
ก.พ.-เม.ย.	0.00	-	-	-	1

\*มีนัยสำคัญทางสถิติที่ระดับ 0.05

หมายเหตุ: Relative Risk (RR) คือความเสี่ยงหรือโอกาสของการป่วยโรคมาลาเรียเปรียบเทียบกับกลุ่มอ้างอิง

ตารางที่ 3 ค่าประมาณ random intercept ของแต่ละจังหวัดในตัวแบบ GLMM

จังหวัด	ค่าประมาณ	ค่าคลาดเคลื่อนมาตรฐาน	ค่าสถิติ t	Pr >  t
เชียงใหม่	-2.17	0.60	-3.64*	0.001
ลำพูน	1.14	0.47	2.43*	0.02
แม่ฮ่องสอน	1.98	0.59	3.32*	0.001
น่าน	-1.47	0.63	-2.31*	0.02
เพชรบูรณ์	-0.93	0.53	-1.78*	0.08
ตาก	1.40	0.59	2.36*	0.02
อุดรดิตถ์	-1.45	0.64	-2.28*	0.02
พิจิตร	-1.16	0.61	-1.91*	0.06
พระนครศรีอยุธยา	-1.04	0.55	-1.89*	0.06
สิงห์บุรี	-1.55	0.84	-1.84*	0.07
ราชบุรี	-1.18	0.59	-1.99*	0.05
สมุทรสาคร	1.96	0.42	4.61*	<.0001
ฉะเชิงเทรา	-1.05	0.53	-1.99*	0.05
นครนายก	-1.88	0.59	-3.22*	0.002
สมุทรปราการ	0.90	0.43	2.11*	0.04
จันทบุรี	1.95	0.51	3.81*	0.0003
ตราด	1.34	0.65	2.07*	0.04
เลย	-1.56	0.73	-2.13*	0.03
มุกดาหาร	1.38	0.53	2.60*	0.01
บุรีรัมย์	-1.74	0.73	-2.39*	0.02
ศรีสะเกษ	1.47	0.69	2.12*	0.04
อุบลราชธานี	2.63	0.49	5.32*	<.0001
ชุมพร	-1.17	0.64	-1.84*	0.07
สุราษฎร์ธานี	1.16	0.47	2.46*	0.02
พัทลุง	-2.84	0.59	-4.80*	<.0001
ตรัง	-2.16	0.53	-4.09*	<.0001
ปัตตานี	1.37	0.47	2.90*	0.005
ยะลา	1.96	0.67	2.93*	0.005
สตูล	-2.43	0.68	-3.57*	0.001

\*มีนัยสำคัญทางสถิติที่ระดับ 0.05

เท่า เมื่อเปรียบเทียบกับผู้ที่อยู่ในจังหวัดไม่ติดชายแดน พบว่าความเสี่ยงของผู้ที่อยู่ในจังหวัดชายแดนติดกับพม่าสูงเป็น 46.77 เท่า ชายแดนติดกับมาเลเซียสูงเป็น 9.73 เท่า และชายแดนติดกับกัมพูชา สูงเป็น 4.56 เท่า เมื่อเปรียบเทียบกับฤดูกาลช่วงเดือน ก.พ.-เม.ย. พบว่าความเสี่ยงในฤดูกาลช่วงเดือน พ.ค.-ก.ค. สูงเป็น 2.14 เท่า และช่วงเดือน พ.ย.-ม.ค. สูงเป็น 0.76 เท่า

ค่าประมาณความแปรปรวนของ random intercept ของแต่ละจังหวัดในตัวแบบ GLMM มีค่า 1.6761 (ค่าความคลาดเคลื่อนมาตรฐาน = 3.166) ส่วนค่าประมาณ random intercept ของแต่ละจังหวัดแสดงในตารางที่ 3 (แสดงเฉพาะจังหวัดที่มีค่าประมาณไม่เท่ากับศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05)

จากตารางที่ 3 พบว่าค่าประมาณของ random intercept ของแต่ละจังหวัดมีค่าแตกต่างกัน เช่น จังหวัดเชียงใหม่ ลำพูน และแม่ฮ่องสอน มีค่า random intercept เท่ากับ -2.17, 1.14 และ 1.98 ตามลำดับ ส่งผลให้การถดถอยของแต่ละจังหวัดมีค่า intercept ต่างกัน เช่น intercept ของจังหวัดเชียงใหม่มีค่าเท่ากับผลรวมของ -13.44 กับ -2.17 คือ -15.61 และ intercept ของจังหวัดลำพูนมีค่าเท่ากับผลรวมของ -13.44 กับ 1.14 คือ -12.30 เป็นต้น แสดงให้เห็นว่าเมื่อไม่มีปัจจัยอื่นใดมาเกี่ยวข้อง อัตราการป่วยโรคมาลาเรียของผู้ที่อยู่ในแต่ละจังหวัดมีค่าแตกต่างกัน ซึ่งเป็นความแตกต่างกันโดยธรรมชาติ (natural heterogeneity) ของแต่ละจังหวัด

### วิจารณ์

ตัวแบบ GEE และ GLMM ที่ตัวแปรตามมีการแจกแจงแบบ Negative Binomial มีความเหมาะสมกว่าแบบ Poisson เนื่องจากข้อมูลผู้ป่วยมาลาเรียที่นำมาศึกษามี over dispersion คือความแปรปรวนมีค่าสูงกว่าค่าเฉลี่ย (ค่าเฉลี่ย = 33.86 8; ค่าความแปรปรวน = 12,600.04) สอดคล้องกับ Hilbe<sup>(19)</sup> ที่เสนอว่าถ้าข้อสมมุติ (assumption) ของ Poisson regression

ที่ว่า ค่าเฉลี่ยเท่ากับค่าความแปรปรวนไม่เป็นจริงแล้ว Negative Binomial regression เป็นทางเลือกที่เหมาะสมเนื่องจากมีข้อจำกัดน้อยกว่า และยังสอดคล้องกับงานวิจัยของ Mwangi และคณะ<sup>(20)</sup> ที่ศึกษาการเกิดโรคมาลาเรียในเด็กอายุต่ำกว่า 15 ปีในเมือง Kilifi ประเทศเคนยา ซึ่งพบว่าข้อมูลที่ศึกษามี over dispersion และตัวแบบ Negative Binomial regression มีความเหมาะสมกว่าตัวแบบ Poisson regression

เมื่อพิจารณาในลักษณะภาพรวมทั้งประเทศจากตัวแบบ GEE พบว่าปัจจัยเสี่ยงของการเกิดโรคมาลาเรียคือ ภาคตะวันออก ภาคใต้ ชายแดนติดกับพม่า ติดกับมาเลเซีย ติดกับกัมพูชา ติดกับลาว ฤดูกาลช่วงเดือน พ.ค.-ก.ค. พ.ย.-ม.ค. ส.ค.-ต.ค. อุณหภูมิเฉลี่ย และปริมาณฝน สอดคล้องกับรายงานของสำนักกระบาดวิทยา<sup>(2)</sup> ที่ว่าแหล่งของเชื้อมาลาเรียอยู่บริเวณป่าเขา ชายแดนของประเทศ โดยเฉพาะบริเวณชายแดนไทย-พม่า และไทย-กัมพูชา และพบผู้ป่วยได้ตลอดทั้งปี แต่จะพบมากในช่วงฤดูฝน สอดคล้องกับงานวิจัยของ Devi และ Jauhari<sup>(21)</sup> ที่ศึกษาความสัมพันธ์ระหว่างสภาพอากาศกับอุบัติการณ์ของโรคมาลาเรีย ในเมือง Dehradun ประเทศอินเดีย ที่พบว่า ปริมาณฝน อุณหภูมิ และความชื้น มีความสัมพันธ์กับอุบัติการณ์ของโรคมาลาเรีย สาเหตุที่ผู้ที่อยู่ในภาคตะวันออก และภาคใต้ มีความเสี่ยงสูงต่อการป่วยโรคมาลาเรีย เนื่องจากลักษณะพื้นที่ส่วนใหญ่ในภาคตะวันออกและภาคใต้ เป็นป่า ภูเขา มีแอ่งน้ำ ลำธาร ซึ่งเป็นแหล่งโรคมาลาเรีย ผู้ที่อยู่ในพื้นที่ในจังหวัดชายแดนมีความเสี่ยงต่อการป่วยโรคมาลาเรียสูงอาจเป็นเพราะว่านอกจากพื้นที่ส่วนใหญ่เป็นป่า ภูเขา มีแอ่งน้ำ ลำธาร แล้วยังมีการติดต่อค้าขายระหว่างประชาชนที่อาศัยอยู่บริเวณชายแดน การจ้างแรงงานต่างชาติ ทำให้เกิดการแพร่เชื้อมาลาเรียได้ง่าย และอีกสาเหตุหนึ่งที่ทำให้ปัจจัยชายแดนภาคใต้มีความเสี่ยงสูงต่อการเกิดโรคมาลาเรีย คือเหตุการณ์ความไม่สงบในพื้นที่ชายแดนภาคใต้ซึ่งทำให้เจ้าหน้าที่สาธารณสุขไม่สามารถเข้าไปควบคุมโรคในพื้นที่ได้



อย่างต่อเนื่อง และฤดูกาลช่วงเดือน พ.ค.-ก.ค. เป็นช่วงที่มีความเสี่ยงสูงเนื่องจากเป็นช่วงที่ฝนตกชุกเหมาะแก่การแพร่พันธุ์ของยุงลายที่เป็นพาหะของโรคมาลาเรีย

### ข้อเสนอแนะในการนำผลงานวิจัยไปใช้

1. หน่วยงานหรือผู้ที่เกี่ยวข้องกับการวางแผนควบคุม และป้องกันโรคมาลาเรียควรจัดโครงการให้ความรู้ วรรณคดีป้องกันโรคมาลาเรียให้กลุ่มที่มีความเสี่ยงสูงก่อน

2. ผู้ที่อยู่ในกลุ่มเสี่ยงควรปฏิบัติตนอย่างเคร่งครัดเพื่อป้องกันการป่วยโรคมาลาเรีย

### ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ใช้ตัวแบบ GEE และ GLMM กับโรคอื่น เช่น โรคไข้เลือดออก

2. ใช้ตัวแบบ GEE และ GLMM กับข้อมูลทางระบาดวิทยาในพื้นที่ขนาดเล็กลง เช่น อำเภอ ตำบล หมู่บ้าน หรือระดับบุคคล

3. เพิ่มเทอมที่เป็นอิทธิพลสุ่ม นอกจาก random intercept เข้าไปในตัวแบบ GLMM เช่น spatial random effect

### กิตติกรรมประกาศ

ขอขอบคุณ สำนักกระบาดวิทยา กรมป่าไม้ กรมอุทยานแห่งชาติ และสำนักงานสถิติแห่งชาติที่เผยแพร่ข้อมูลที่ใช้วิจัย

### เอกสารอ้างอิง

1. WHO. Malaria 2008. Fact sheet N° 94. [online] 2008 [cited 2009 Jan 10]; Available from: URL: <http://www.who.int/mediacentre/factsheets/fs094/en/>
2. สำนักกระบาดวิทยา. โรคมาลาเรีย. สรุปรายงานการเฝ้าระวังโรค 2550 [online] 2551 [สืบค้นเมื่อ 20 มกราคม 2552]; Available from: URL: [http://203.157.15.4/Annual/AN-NUAL2550/Part1/0550\\_Malaria.doc](http://203.157.15.4/Annual/AN-NUAL2550/Part1/0550_Malaria.doc)

3. Yeshiwondim AK, Gopal S, Hailemariam AT, Dengela DO, Patel HP. Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. *International Journal of Health Geographics* 2009; 8:5.
4. Tian L, Bi Y, Ho SC, Liu W, Liang S, Boggins WE, et al. One-year delayed effect of fog on malaria transmission: a time-series analysis in the rain forest area of Mengla County, south-west China. *Malaria Journal* 2008; 7:110-8.
5. รัศมี ศรีชื่น. ปัจจัยเสี่ยงของการติดเชื้อมาลาเรียในพื้นที่ชายแดนไทย- สหภาพเมียนมาร์ จังหวัดระนอง (วิทยา-นิพนธ์ปริญญาสาธารณสุขศาสตรมหาบัณฑิต). ภาควิชาชีวสถิติและประชากรศาสตร์, บัณฑิตวิทยาลัย. ขอนแก่น: มหาวิทยาลัยขอนแก่น; 2548.
6. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73:13-22.
7. Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz MA. Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology*; 1998; 147(7):694-703.
8. Yesilova A, Yilmaz A. The application of overdispersion and generalized estimating equations in repeated categorical data related to the sexual behaviour traits of farm animals. *Journal of Applied Sciences* 2007; 7(12):1762-7.
9. สำนักกระบาดวิทยา. จำนวนผู้ป่วย-เสียชีวิตรายเดือน ปี 2550 มาลาเรีย. สรุปรายงานการเฝ้าระวังปี [online] [สืบค้นเมื่อ 20 มกราคม 2552]; Available from: URL: [http://203.157.15.4/surdata/y50/mcd\\_Malaria\\_50.rtf](http://203.157.15.4/surdata/y50/mcd_Malaria_50.rtf)
10. กรมป่าไม้. เนื้อที่ป่าไม้ของประเทศไทย [online] 2551 [สืบค้นเมื่อ 20 มกราคม 2552]; Available from: URL: <http://www.forest.go.th/home/index.asp>
11. กรมอุทยานแห่งชาติ. ปริมาณฝนและอุณหภูมิ ปี 2550 [online] 2551 [สืบค้นเมื่อ 20 มกราคม 2552]; Available from: URL: <http://www.tmd.go.th/index.php>
12. สำนักงานสถิติแห่งชาติ. รายได้เฉลี่ยต่อหัวต่อปี ปี 2550 [online] 2551 [สืบค้นเมื่อ 20 มกราคม 2552]; Available from: URL: [http://service.nso.go.th/nso/nso\\_center/project/search\\_center/23project-th.htm](http://service.nso.go.th/nso/nso_center/project/search_center/23project-th.htm)
13. Hojsgaard S. Generalized estimating equations (GEE) for glm-type data. Danish Institute of Agricultural Sciences [online] 2006 [cited 2009 Jan 10]; Available from: URL: <http://staff.pubhealth.ku.dk/~pd/mixed-jan.2006/R-mixed-geeglm-Lecture.pdf>
14. Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. *The American Statistician* 1999; 53:160-9.
15. Ballinger GE. Using generalized estimating equations

- for longitudinal data analysis. *Organizational Research Methods* 2004; 7(2):127-50.
16. Ho KT, Ahn CW, Alarcon GS, Baetjge BA, Tan FK, Roseman J, et al. Systemic lupus erythematosus in multiethnic cohort (LUMINA): XXVIII. Factors predictive of thrombotic events. *Rheumatology* 2005; 44: 1303-7.
  17. UCLA Academic Technology Services. SAS Annotated output : negative binomial regression [online] [cited 2009 August 8]; Available from: URL: UCLA Academic Technology Services.
  18. Schabenberger O. Introducing the GLIMMIX procedure for generalized linear mixed models [online] [cited 2009 August 8]; Available from: URL: <http://www2.sas.com/proceedings/sugi30/196-30.pdf>
  19. Hilbe JM. *Negative binomial regression*. New York: Cambridge University Press; 2007.
  20. Mwangi TW, Fegan G, Williams TN, Kinyanjui SM, Snow RW, Marsh K. Evidence for over-dispersion in the distribution of chincinal malaria episodes in children. *PLoS ONE* 2008; 3(5): e2196.
  21. Devi NP, Jauhari RK. Climatic variables and malaria incidence in Dehradun, Uttaranchal, India. *J Vector Borne Dis* 2006; 43(1):21-8.

**Abstract Risk Factors for Malaria in Thailand Using Generalized Estimating Equation (GEE) and Generalized Linear Mixed Model (GLMM)**

**Krisada Lekdee, Lily Ingsrisawang**

Department of Statistics, Faculty of Science, Kasetsart University

*Journal of Health Science* 2010; 19:364-73.

The objective of this research involved the identification of risk factors for malaria in Thailand using suitable statistical models. The Generalized Estimating Equation (GEE) and the Generalized Linear Mixed Model (GLMM) were employed. For both the GEE and the GLMM, the dependent variable with a Poisson distribution was compared with a Negative Binomial distribution. Secondary and provincial-level data used in this research were from several sources. The dependent variable was the number of people with malaria in 2007, and the independent variables were region, border, season, forest area, rain, temperature, and income. Suitable models were judged by the value of mean deviance (for GEE) and mean generalized chi-square (for GLMM). The study found that for both the GEE and the GLMM, the dependent variable with Negative Binomial distribution was more suitable than was the Poisson distribution. For the GEE, the population-averaged model, region (East and South), border (Myanmar, Malaysia, Cambodia, and Laos), season (May-Jul, Nov-Jan, and Aug-Oct), temperature, and rain significantly impacted the malaria incidence rate. For the GLMM, subject-specific model, factors that significantly related to the malaria incidence rate were region (East, South, and North-East), border (Myanmar, Malaysia, and Cambodia), and season (May-Jul and Nov-Jan).

**Key words:** malaria, Generalized Estimating Equation (GEE), Generalized Linear Mixed Model (GLMM), Poisson, Negative Binomial