

การวิเคราะห์ปัจจัยที่ส่งผลต่อการมีชีวิตรอด และสถานะของผู้ป่วยโรคเอดส์ด้วยตัวแบบเชิงเส้นนัยทั่วไป An Analysis of Factors Effecting the Status of Living and Karnofsky Performance Scales of Aids Patients Using Generalized Linear Models

บุญนุษ ัญญกุลสัจจา, ลลิตา โพธิ์ทอง และวีรานันท์ พงศาภักดี*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์

ตำบลสนามจันทร์ อำเภอเมือง จังหวัดนครปฐม 73000

Punyanut Tanyakunsadcha, Lalita Phothong and Veeranun Pongsapukdee*

Department of Statistics, Faculty of Science, Silpakorn University, Sanam Chandra Palace Campus,
Sanam Chandra, Muang, Nakhon Pathom 73000

บทคัดย่อ

การวิเคราะห์การมีชีวิตรอดและสถานะกิจกรรมของผู้ป่วยโรคเอดส์ด้วยตัวแบบเชิงเส้นนัยทั่วไปแบ่งเป็น 2 กรณี คือ ตัวแปรตอบสนองมี 2 กลุ่ม : มีชีวิตรอดและเสียชีวิต อาศัยตัวแบบลอจิสต์ 2 กลุ่ม (binary logit model) และกรณีที่ตัวแปรตอบสนองมี 3 กลุ่ม : สถานะที่ผู้ป่วยสามารถทำกิจกรรมได้ตามปกติ ผู้ป่วยสามารถทำกิจกรรมได้ตามปกติแต่เริ่มมีอาการแสดงของโรค และผู้ป่วยไม่สามารถทำกิจกรรมได้ตามปกติ อาศัยตัวแบบลอจิสต์สะสม (cumulative logit model) โดยศึกษาปัจจัยที่เกี่ยวข้องจำนวน 11 ตัวแปร ได้แก่ ระยะเวลาตั้งแต่วินิจฉัยการเป็นโรคเอดส์จนกระทั่งเสียชีวิตหรือสิ้นสุดการทดลอง ผลการวินิจฉัยระยะเวลากการมีชีวิตรอดของผู้ป่วย ประเภทของการรักษาผู้ป่วย ประเภทกลุ่มยาที่ใช้ในการรักษาผู้ป่วย เพศ เชื้อชาติ ประวัติการใช้ยาเสพติด โรคฮีโมฟีเลีย ระดับเซลล์เม็ดเลือดขาว ระยะเวลาของการใช้ยา zidovudine และอายุ จากข้อมูลจริงจำนวน 1,151 คน การประมวลผลใช้โปรแกรมสำเร็จรูป SAS Enterprise Guide version 5.1 ผลการวิจัยพบว่าปัจจัยที่ส่งผลต่อการมีชีวิตรอดของผู้ป่วยโรคเอดส์อย่างมีนัยสำคัญทางสถิติที่ 0.05 มีจำนวน 3 ปัจจัย คือ ระดับเซลล์เม็ดเลือดขาว ประเภทของการรักษาผู้ป่วย และผลการวินิจฉัยระยะเวลากการมีชีวิตรอดของผู้ป่วย ส่วนปัจจัยที่ส่งผลต่อสถานะกิจกรรมของผู้ป่วยอย่างมีนัยสำคัญทางสถิติที่ 0.05 คือ ระยะเวลาตั้งแต่วินิจฉัยการเป็นโรคเอดส์จนกระทั่งเสียชีวิตหรือสิ้นสุดการทดลอง ระดับเซลล์เม็ดเลือดขาว และอายุ โดยตัวแบบเชิงสถิติทั้ง 2 กรณี มีภาวะสารูปดีและเหมาะสมกับข้อมูลอย่างมีนัยสำคัญทางสถิติที่ 0.05

คำสำคัญ : ตัวแบบเชิงสถิติ; การพยากรณ์ความน่าจะเป็น; ตัวแบบลอจิสต์สองกลุ่ม; ตัวแบบลอจิสต์สะสม

*ผู้รับผิดชอบบทความ : veeranun@su.ac.th

Abstract

The analysis of status of living and kanofsky performance scale of aids patients consists of two cases. Firstly, the response was divided into two groups: alive and death, using a binary logit model. Secondly, the response was divided into three groups: normal activities, normal activities with the sense of disease, and unable to do activities, using a cumulative logit model. The eleven factors consist of time from AIDS diagnosis to death, estimated time to death, treatment indicator, treatment group indicator, sex, race, drug use history, hemophiliac, white blood cell level, prior ZDV use, and age. The real experimental data were from 1,151 AIDS patients. All data was processed using SAS Enterprise Guide version 5.1. The results revealed that the factors effecting the status living of the response significantly at 0.05 are white blood cell level (cd4), treatment indicator, and estimated time to death. The factors effecting the kanofsky performance scale of the response significantly at 0.05 are time from AIDS diagnosis to death (time), cd4, and age. Moreover, both logit models have adequacy of fits for the data significantly at 0.05.

Keywords: statistical model; probability forecasting; logit model; binary logit mode; cumulative logit model

1. บทนำ

โรคเอดส์เป็นที่รู้จักกันในวงกว้างและมีผู้เสียชีวิตเกือบ 30 ล้านคนทั่วโลก เนื่องจากเป็นโรคติดต่อที่ร้ายแรง ซึ่งสาเหตุของโรคเกิดจากไวรัสเข้าสู่ร่างกาย เข้าไปติดเชื้อและเพิ่มจำนวนในเซลล์เม็ดเลือดขาว ทำให้เซลล์เม็ดเลือดขาวถูกทำลายและลดจำนวนลงเรื่อย ๆ จนในที่สุดระบบภูมิคุ้มกันของร่างกายก็ไม่สามารถทำหน้าที่ต่อสู้ทำลายเชื้อโรคได้ ในขณะนี้ยังไม่มียาตัวไหนที่สามารถรักษาโรคเอดส์ให้หายขาดได้ มีเพียงยาที่ใช้เพื่อช่วยยับยั้งไม่ให้ไวรัสเอชไอวีเพิ่มจำนวนมากขึ้นในร่างกายเท่านั้น อย่างไรก็ตามในปัจจุบันอัตราของผู้ป่วยโรคเอดส์รายใหม่ทั่วโลกได้ลดลงเกือบร้อยละ 25 แต่จำนวนผู้ป่วยที่เป็นโรคเอดส์ทั่วโลกยังคงมีจำนวนที่ค่อนข้างสูงมาก [1]

ความสำคัญของการดูแลและรักษาผู้ป่วยโรคเอดส์ ซึ่งในปัจจุบันมีผู้ป่วยโรคเอดส์ที่เสียชีวิตอย่างรวดเร็วและที่ยังมีชีวิตอยู่และรอรับการรักษายังเป็น

จำนวนมาก [2] ทำให้ผู้ป่วยตระหนักถึงสภาวะการมีชีวิตอยู่ของผู้ป่วยที่สามารถมีชีวิตยืนยาวไปอีกระยะหนึ่ง และสภาวะที่ผู้ป่วยสามารถทำกิจกรรมได้ตามปกติ หรือผู้ป่วยสามารถทำกิจกรรมได้ตามปกติแต่เริ่มมีอาการแสดงของโรค และผู้ป่วยที่ไม่สามารถทำกิจกรรมได้ โดยทำการศึกษาปัจจัยที่เกี่ยวข้องด้านต่าง ๆ ที่อาจส่งผลกระทบต่อเชิงสถิติต่อกลุ่มของผู้ป่วยดังกล่าว

การวิเคราะห์และการศึกษาปัจจัยที่ส่งผลต่อการมีชีวิตอยู่และการประเมินสภาวะผู้ป่วยโรคเอดส์ อาจทดสอบความเกี่ยวพันระหว่างปัจจัยที่ส่งผลต่อการมีชีวิตอยู่ของผู้ป่วยโรคเอดส์และที่ส่งผลต่อการประเมินสภาวะผู้ป่วยของผู้ป่วยโรคเอดส์ รวมทั้งการใช้ตัวแบบที่เหมาะสมในการพยากรณ์กลุ่มที่สนใจของผู้ป่วยโรคเอดส์ เพื่อหาสารสนเทศที่อาจใช้ประโยชน์ควบคู่กับการดูแลผู้ป่วย ให้สามารถมีชีวิตยืนยาวไปอีกระยะหนึ่ง และมีสภาวะที่ผู้ป่วยสามารถทำกิจกรรมได้

ตามปกติ งานวิจัยนี้จึงทำการวิเคราะห์ข้อมูลภายใต้ตัวแบบลอจิสติกสองกลุ่ม (binary logit model) และตัวแบบลอจิสติกสะสม (cumulative logit model) ที่มีการประเมินว่ามีภาวะสารูปดีและเหมาะสมกับข้อมูลด้วย โดยศึกษาภายใต้ข้อมูลจริงจากงานสำรวจทดลองทางสาธารณสุข ซึ่งสำรวจผู้ป่วยโรคเอดส์จำนวน 1,151 คน [3] และกำหนดขอบเขตของตัวแปรดังนี้ ตัวแปรตอบสนอง (Y) แบ่งเป็น 2 กรณี คือ

กรณีที่ 1 การใช้ตัวแบบลอจิสติกสองกลุ่ม (dichotomous logit model, [4]) เมื่อตัวแปรตอบสนอง (Y) มี 2 กลุ่ม ได้แก่ มีชีวิตอยู่ และเสียชีวิต

กรณีที่ 2 การใช้ตัวแบบลอจิสติกสะสม [5-7,9] เมื่อมีตัวแปรตอบสนอง (Y) คือ การประเมินสภาวะผู้ป่วย (Y) 3 กลุ่ม ได้แก่ ผู้ป่วยสามารถทำกิจกรรมได้ตามปกติ ผู้ป่วยสามารถทำกิจกรรมได้ตามปกติแต่เริ่มมีอาการแสดงของโรค และผู้ป่วยไม่สามารถทำกิจกรรมได้ตามปกติ ส่วนปัจจัยหรือตัวแปรอธิบาย (X's) ที่เกี่ยวข้องและนำมาศึกษาทั้งกรณี 1-2 ศึกษาจากปัจจัยจำนวน 11 ปัจจัย ดังนี้

1. ระยะเวลาตั้งแต่วินิจฉัยการเป็นโรคเอดส์จนกระทั่งเสียชีวิตหรือสิ้นสุดการทดลอง (time) ได้แก่ ภายใน 12 เดือน ภายใน 9 เดือน ภายใน 6 เดือน และภายใน 3 เดือน

2. ผลการวินิจฉัยระยะเวลาการมีชีวิตอยู่ของผู้ป่วย (dtime) ได้แก่ ภายใน 12 เดือน ภายใน 9 เดือน ภายใน 6 เดือน และภายใน 3 เดือน

3. ประเภทของการรักษาผู้ป่วย (tx) ได้แก่ กลุ่มการรักษาโดยให้ยา indinavir (IDV) และกลุ่มควบคุม (กลุ่มการรักษาโดยไม่มีการให้ยา indinavir)

4. ประเภทกลุ่มยาที่ใช้ในการรักษาผู้ป่วย (txgrp) ได้แก่ (1) ZDV และ 3TC (2) ZDV กับ 3TC และ IDV (3) d4T และ 3TC และ (4) d4T กับ 3TC และ IDV

5. เพศ (sex) ได้แก่ เพศชาย และเพศหญิง

6. เชื้อชาติหรือลักษณะของเชื้อชาติ (race) ได้แก่ ชาวอเมริกันมีเชื้อสายฮิสแปนิก ชาวอเมริกันผิวขาวไม่มีเชื้อสายฮิสแปนิก ชาวอเมริกันผิวดำไม่มีเชื้อสายฮิสแปนิก ชาวเอเชีย ชาวอเมริกันมีเชื้อสายอินเดีย

7. ประวัติการใช้ยาเสพติด (ivdrug) ได้แก่ ไม่เคยใช้ยาเสพติด เคยใช้ยาเสพติดมาแล้วก่อนหน้านี้ และใช้ยาเสพติดอยู่

8. โรคฮีโมฟีเลีย (hemophilia) ได้แก่ ไม่เป็นโรคฮีโมฟีเลีย และเป็นโรคฮีโมฟีเลีย

9. ระดับเซลล์เม็ดเลือดขาว (cd4) ได้แก่ ระดับสูงกว่า 50 และระดับต่ำกว่าหรือเท่ากับ 50

10. ระยะเวลาของการให้ยา zidovudine มาก่อน (priorzdv) ได้แก่ มากกว่า 5 ปี ภายใน 5 ปี ภายใน 4 ปี ภายใน 3 ปี ภายใน 2 ปี และภายใน 1 ปี

11. อายุ (age) ได้แก่ 12-18 ปี : วัยรุ่น 19-35 ปี : วัยผู้ใหญ่ตอนต้น 36-60 ปี : วัยผู้ใหญ่ตอนกลาง และมากกว่า 60 ปี : วัยผู้ใหญ่ตอนปลายหรือวัยชรา

การประมวลผลข้อมูลอาศัยโปรแกรมสำเร็จรูป SAS Enterprise Guide version 5.1 ลิขสิทธิ์ของภาคอุตสาหกรรม คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร 2556 [8]

2. วิธีการทดลองและขอบเขตของการวิจัย

ตัวแบบเชิงสถิติเพื่อการวิเคราะห์ข้อมูลอาศัยตัวแบบเชิงเส้นน้อยทั่วไป (generalized linear model) 2 ตัวแบบ ดังนี้

2.1 ตัวแบบลอจิสติก 2 กลุ่ม [4]

เมื่อตัวแปรอธิบายหรือปัจจัยมีหลายตัว (x หรือ X 's) คือ $Q(x) = \log\left[\frac{P(x)}{1-P(x)}\right] = \log\left(\frac{P(Y=1|x)}{1-P(Y=1|x)}\right)$
 $= \beta_0 + \beta'x$... (1) เมื่อ β_0 , β คือ จุดตัด และเวกเตอร์ความชัน ตามลำดับ

2.2 ตัวแบบลอจิสติกส์ [5]

ใช้กรณีตัวแบบ The proportional odds logit models [6] เมื่อมีตัวแปรตอบสนองแบบอันดับ Y มี 3 กลุ่ม และตัวแปรอธิบาย (x หรือ x 's) โดยเรียกตัวแบบนี้ชื่อหนึ่งว่าตัวแบบลอจิสติก 3 กลุ่ม (trichotomous logit model) [6,7] คือ $\log \left(\frac{P(Y \leq j|x)}{1 - [P(Y \leq j|x)]} \right)$

$$= \log \left(\frac{P_1(x) + \dots + P_j(x)}{P_{j+1} + \dots + P_J(x)} \right) = \text{logit} [P(Y \leq j|x)] = \beta_{j0} + \beta'x$$

...(2) เมื่อ β_{j0} , β คือ จุดตัด และเวกเตอร์ความชันตามลำดับ โดยที่ $j = 1, \dots, J-1$.

2.3 วิธีการวิเคราะห์ข้อมูล

แบ่งข้อมูลทั้งหมดออกเป็น 2 ส่วน ส่วนแรก 80 % (training set) คือ ผู้ป่วยจำนวน 921 คน ซึ่งใช้สำหรับการทดลองทุกขั้นตอน และส่วนหลัง 20 % (validation set) คือ ผู้ป่วยจำนวน 231 คนซึ่งใช้สำหรับการตรวจสอบความน่าเชื่อถือของตัวแบบที่สร้างจากข้อมูลในส่วนแรก ขั้นตอนต่อไปคือการสร้างตัวแบบ (1) และ (2) ด้วย SAS และใช้วิธีคัดเลือกแบบขั้นตอน (stepwise selection method) โดยทำการวิเคราะห์ภาวะสารถูปดีของตัวแบบและการตรวจสอบความเหมาะสมของตัวแบบด้วยเกณฑ์อื่น ๆ การทดสอบปัจจัยต่าง ๆ และตรวจสอบความสัมพันธ์ด้วยมาตรวัดความสอดคล้องระหว่างค่าสังเกตกับค่าความน่าจะเป็นของการพยากรณ์ เพื่อการพยากรณ์ค่าความน่าจะเป็นต่อไป รวมทั้งการสร้างตารางการจำแนกกลุ่ม (classification table) เพื่อดูความถูกต้องในการพยากรณ์ของตัวแบบที่ใช้ [9]

2.4 การทดสอบพารามิเตอร์

การทดสอบความมีนัยสำคัญของพารามิเตอร์ที่แสดงอิทธิพลของตัวแปรอธิบายต่อตัวแปรตอบสนอง โดยผลลัพธ์จาก SAS ให้ผลการทดสอบต่อไปนี้ (1) ตัวสถิติทดสอบ Wald test (2) ตัวสถิติ G^2

หรือ Deviance หรือ ตัวสถิติอัตราส่วนความน่าจะเป็นไคสแควร์ (likelihood-ratio chi-squared statistic) และ (3) การทดสอบสกอร์ (score test)

2.5 การทดสอบความเหมาะสมของตัวแบบ

การทดสอบภาวะสารถูปดีของตัวแบบ อาจใช้ตัวสถิติเพียร์สันไคสแควร์ (Pearson chi-squared statistic) และตัวสถิติอัตราส่วนความน่าจะเป็นไคสแควร์ (likelihood-ratio chi-squared statistic) การพยากรณ์ค่าความน่าจะเป็นจะอาศัยตัวแบบที่เหมาะสมที่ได้ทดสอบและตรวจสอบความเหมาะสมเรียบร้อยแล้ว

2.6 มาตรวัดความสอดคล้อง

นอกจากนี้อาจตรวจสอบตัวแบบด้วยขนาดความเกี่ยวพันหรือมาตรวัดความสอดคล้องระหว่างค่าสังเกตกับค่าจากตัวแบบหรือค่าพยากรณ์ ได้แก่ พิจารณาจากตัวสถิติ ดังนี้ (1) ตัวสถิติ gamma (Goodman and Kruskal's, 1954) (2) ตัวสถิติ Somers'D (3) ค่าสถิติ C และ (4) ตัวสถิติ Kendall's Tau

2.6 การพยากรณ์

ตัวสถิติที่อาจใช้ตรวจสอบประสิทธิภาพในการพยากรณ์โดยทั่วไปนิยมใช้ตัวสถิติ sensitivity, specificity [5,9] เมื่อ sensitivity หรือความไวของตัวแบบวัดด้วย สัดส่วนของการพยากรณ์ถูกในกลุ่มที่สนใจ หาดด้วยจำนวนในกลุ่มที่สนใจ ส่วน specificity คือ สัดส่วนของการพยากรณ์ถูกในกลุ่มที่ไม่สนใจ หาดด้วยจำนวนในกลุ่มที่ไม่สนใจ นอกจากนี้ประสิทธิภาพในการพยากรณ์ยังนิยมตรวจสอบจากเส้นโค้ง ROC (receiver-operating characteristic curve) โดยการวิเคราะห์ความไวของตัวแบบด้วย ROC curve มีแกนตั้งแสดงค่า sensitivity และแกนนอนแสดงค่า 1-specificity ถ้าจุดตัดน้อยไปหรือมากไป จะมีผลต่อ sensitivity และ specificity ถ้า sensitivity มากขึ้น

specificity จะลดลง และถ้า sensitivity ลดลง specificity จะเพิ่มขึ้น นอกจากนี้พื้นที่ภายใต้ ROC curve ยังใช้เป็นตัววัดความสอดคล้องระหว่างค่าสังเกตกับค่าความน่าจะเป็นของการพยากรณ์ ที่เรียกว่า ดัชนีความสอดคล้อง (concordant index) หรือตัวสถิติ c เส้นในแนวตั้งที่ต้องการคือเส้นที่ให้ curve ห่างจากเส้นทแยงมุมไปใกล้กับแกนตั้งมากที่สุด (distance to be maximized) นั่นคือจุดที่แสดงว่าตัวแบบเหมาะสมมากหรือเส้นโค้งมีความชันสูง คือ มีจุดภายในเส้นโค้งที่อยู่ใกล้จุด sensitivity = 100 % มากที่สุด [5,9] (รูปที่ 1)

3. ผลการวิจัย

การวิเคราะห์ข้อมูลด้วยตัวแบบลอจิสต์ 2 กลุ่ม หรือกรณีตัวแปรตอบสนองจำแนกประเภท 2 กลุ่ม ได้แก่ การมีชีวิตอยู่และเสียชีวิตของผู้ป่วย พบว่าผลการทดสอบพารามิเตอร์ของสัมประสิทธิ์ปัจจัยของตัวแบบด้วยตัวสถิติ Wald มีความเหมาะสมที่จะใช้ในการพยากรณ์ (P-value < 0.05) ภายใต้ระดับนัยสำคัญทางสถิติ 0.05 ประกอบด้วย ตัวแปรระยะเวลาการมีชีวิตอยู่ของผู้ป่วย (dtime) ตัวแปรประเภทของการรักษาผู้ป่วย (tx) และตัวแปรระดับเซลล์เม็ดเลือดขาว (cd4) (ตารางที่ 1)

ตารางที่ 1 ค่าสถิติ Wald และ P-value สำหรับปัจจัยที่มีผลต่อตัวแปรตอบสนอง (training set)

อิทธิพล	องศาอิสระ	ค่าสถิติ Wald	P-value
dtime	3	11.2952	0.0102
tx	3	5.4749	0.0193
cd4	1	6.7490	0.0094

เมื่อตรวจสอบตัวแบบลอจิสต์ 2 กลุ่ม ที่อธิบายด้วยปัจจัย ระยะเวลาการมีชีวิตอยู่ของผู้ป่วย (dtime) ประเภทของการรักษาผู้ป่วย (tx) และระดับเซลล์เม็ดเลือดขาว (cd4) ด้วยตัวสถิติ Likelihood ratio, Score และ Wald พบว่าตัวแบบมีภาวะสาธูปดีและเหมาะสมกับข้อมูลโดยตัวแปรปัจจัยที่ใช้ในการพยากรณ์ตัวแปรตอบสนอง 2 กลุ่ม แบบนั้นมีความสำคัญทางสถิติ 0.05 (ตารางที่ 2)

ตารางที่ 2 ตัวสถิติทดสอบและผลการทดสอบ (training set) ตัวแบบลอจิสต์ 2 กลุ่ม

ตัวสถิติทดสอบ	Chi-Square	DF	P-value
Likelihood Ratio	35.0817	5	< 0.0001
Score	32.7129	5	< 0.0001
Wald	22.3611	5	0.0004

นอกจากนี้สามารถหาค่าประมาณสัมประสิทธิ์ของปัจจัยจำแนกตามรายการกลุ่มของปัจจัย เพื่อนำมาสร้างตัวแบบจากข้อมูลส่วนแรก (training set) พร้อมการทดสอบเพิ่มเติม พบว่าสัมประสิทธิ์ของตัวแบบลอจิสต์ของตัวแปรปัจจัยเชิงกลุ่มทั้ง 3 ตัว ได้แก่ dtime, tx และ cd4 ต่างมีนัยสำคัญทางสถิติที่ 0.05 (training set) ทุกตัวแปร (P-value < 0.05) (ตารางที่ 3)

โดยสามารถสร้างสมการลอจิสต์ 2 กลุ่ม ดัง (1.1) คือ $\text{logit}[P(Y=1|x)] = 4.0388 + 2.3187 \text{ dtime} (1) - 0.2178 \text{ dtime} (2) - 1.0313 \text{ dtime} (3) + 0.6671 \text{ tx} (1) + 0.6368 \text{ cd4} (1) \dots (1.1)$

ผลการทวนสอบประสิทธิภาพในการพยากรณ์ของตัวแบบลอจิสต์ 2 กลุ่ม ด้วยข้อมูลส่วนหลัง (validation set) ที่ใช้สำหรับตรวจสอบความถูกต้องของตัวแบบ (1.1) แสดงไว้ในตารางที่ 4

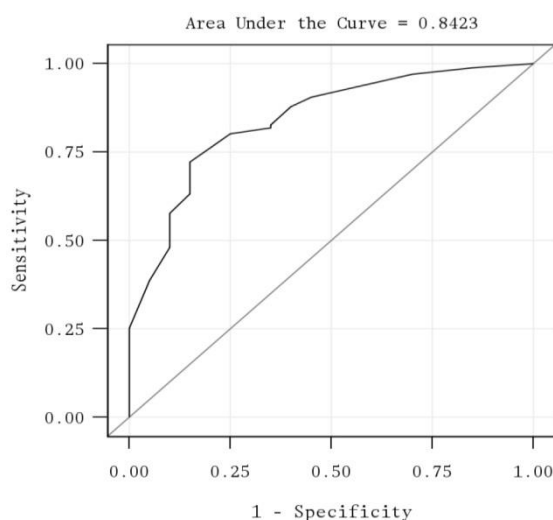
ตารางที่ 3 ค่าประมาณพารามิเตอร์สำหรับตัวแบบลอจิสต์ 2 กลุ่ม

อิทธิพล	กลุ่ม	องศาอิสระ	β	S.E. β	ค่าสถิติ Wald	P-value
Intercept		1	4.0388	0.3599	125.9647	< 0.0001
dtime	1	1	2.3187	0.7756	8.9378	0.0028
dtime	2	1	- 0.2178	0.4072	0.2860	0.5928
dtime	3	1	- 1.0313	0.4232	5.9380	0.0148
tx	1	1	0.6671	0.2851	5.4749	0.0193
cd4	1	1	0.6368	0.2451	6.7490	0.0094

ตารางที่ 4 การจำแนกผลการพยากรณ์ของตัวแบบลอจิสต์ 2 กลุ่ม (validation set)

ค่าสังเกต	ค่าพยากรณ์		
	มีชีวิตอยู่	มีชีวิตอยู่	ร้อยละ
มีชีวิตอยู่	225	0	100.00
เสียชีวิต	3	3	50.00
ภาพรวม			98.70

จากตารางที่ 4 พบว่าการพยากรณ์ด้วยตัวแบบลอจิสต์ 2 กลุ่ม (1.1) ในผู้ป่วยโรคเอดส์ที่มีชีวิตอยู่สามารถพยากรณ์ได้ถูกต้องร้อยละ 100.00 และถ้าผู้ป่วยโรคเอดส์เสียชีวิตจะพยากรณ์ได้ถูกต้องร้อยละ 50.00 ดังนั้นความถูกต้องในการพยากรณ์ของตัวแบบลอจิสต์ 2 กลุ่ม ในภาพรวมสามารถพยากรณ์ได้ถูกต้องร้อยละ 98.70 (รูปที่ 1) และผลการพยากรณ์ด้วยข้อมูลจาก Validation set ใกล้เคียงกับข้อมูล training set ภายใต้วแบบเดียวกัน ดังนั้นตัวแบบลอจิสต์ 2 กลุ่ม จึงมีประสิทธิภาพและความเหมาะสมกับข้อมูลในการนำไปวิเคราะห์ปัจจัยที่ส่งผลต่อตัวแปรตอบสนอง และอาจนำไปใช้พยากรณ์ความน่าจะเป็นที่สนใจหรือกลุ่มที่สนใจของตัวแปรตอบสนองข้างต้นต่อไป โดยอาจตรวจสอบเพิ่มเติมด้วย ROC curve ดังรูปที่ 1



รูปที่ 1 ROC curve ของการพยากรณ์ด้วยตัวแบบลอจิสต์ 2 กลุ่ม

จากรูปที่ 1 เมื่อพิจารณา ROC curve พบว่า curve อยู่ห่างจากเส้นทแยงมุมมากและมีค่าเข้าใกล้ sensitivity 100 % หรือ curve มีความชันสูง ทำให้มีพื้นที่ใต้เส้นโค้งมาก (0.8423) ซึ่งแสดงการสนับสนุนความเหมาะสมของตัวแบบ หมายความว่า การวิเคราะห์และการพยากรณ์ด้วยตัวแบบนี้มีความถูกต้องในภาพรวมค่อนข้างสูง

ส่วนการวิเคราะห์ข้อมูลด้วยตัวแบบลอจิสต์ผสม ในกรณีตัวแปรตอบสนองสภาวะกิจกรรมของผู้ป่วย 3 กลุ่ม นั่นคือ ผู้ป่วยสามารถทำกิจกรรมได้

ตามปกติ ผู้ป่วยสามารถทำกิจกรรมได้ตามปกติแต่เริ่มมีอาการแสดงของโรค และผู้ป่วยไม่สามารถทำกิจกรรมได้ตามปกติ ผลการทดสอบภาวะสารูปดีของตัวแบบหลังนี้ พบว่าตัวแบบมีภาวะสารูปดี มีความเหมาะสมที่จะใช้ในการพยากรณ์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05 โดยตัวแบบลอจิสติกมีความเหมาะสมและสอดคล้องกับการทดสอบพารามิเตอร์ของตัวแบบ ด้วยตัวสถิติ χ^2 Likelihood ratio, score และ Wald ภายใต้ระดับนัยสำคัญทางสถิติ 0.05 ดังผลลัพธ์ในตารางที่ 5-6

ตารางที่ 5 ค่าสถิติ Wald และ P-value สำหรับแต่ละปัจจัยที่มีผลต่อตัวแปรตอบสนอง

อิทธิพล	องศาอิสระ	ค่าสถิติ Wald	P-value
time	3	10.7038	0.0134
cd4	1	31.4268	< 0.0001
age	3	13.9346	0.0030

จากตารางที่ 5 พบว่าปัจจัยที่มีผลต่อการประเมินสภาวะผู้ป่วยโรคเอดส์ ได้แก่ ระยะเวลาตั้งแต่

วินิจฉัยการเป็นโรคเอดส์จนกระทั่งเสียชีวิตหรือสิ้นสุดการทดลอง (time) ระดับเซลล์เม็ดเลือดขาว (cd4) และอายุ (age) อย่างมีนัยสำคัญทางสถิติ 0.05

ตารางที่ 6 ตัวสถิติทดสอบและผลการทดสอบตัวแบบลอจิสติก (training set)

ตัวสถิติทดสอบ	Chi-Square	DF	P-value
Likelihood ratio	53.3517	7	< 0.0001
Score	52.3044	7	< 0.0001
Wald	51.1279	7	< 0.0001

จากตารางที่ 6 สามารถสรุปว่าตัวแบบมีภาวะสารูปดีและปัจจัยหรือตัวแปรจากการวิเคราะห์ด้วยตัวแบบลอจิสติก โดยใช้วิธีการเลือกแบบขั้นตอนประกอบไปด้วยตัวแปรระยะเวลาดั้งแต่วินิจฉัยการเป็นโรคเอดส์จนกระทั่งเสียชีวิตหรือสิ้นสุดการทดลอง (time) ตัวแปรระดับเซลล์เม็ดเลือดขาว (cd4) และตัวแปรอายุ (age) นั้นมีความเหมาะสมที่จะใช้ในการพยากรณ์ตัวแปรตอบสนองจำแนกประเภท 3 กลุ่มแบบมีลำดับ อย่างมีนัยสำคัญทางสถิติ 0.05

ตารางที่ 7 ค่าประมาณพารามิเตอร์สำหรับตัวแบบลอจิสติก 3 กลุ่ม

อิทธิพล	กลุ่ม	องศาอิสระ	β	S.E. β	ค่าสถิติ Wald	P-value
Intercept	1	1	- 1.2536	0.2511	24.9234	< 0.0001
Intercept	2	1	0.8876	0.2492	12.6887	0.0004
time	1	1	- 0.00588	0.0988	0.0035	0.9525
time	2	1	0.3352	0.1062	9.9576	0.0016
time	3	1	- 0.2521	0.1374	3.3686	0.0665
cd4	1	1	0.3672	0.0655	31.4268	< 0.0001
age	1	1	- 0.9186	0.4215	4.7493	0.0293
age	2	1	0.3105	0.2522	1.5162	0.2182
age	3	1	0.6248	0.2555	5.9799	0.0145

จากตารางที่ 7 พบว่าค่าประมาณสัมประสิทธิ์ของตัวแบบลอจิสติกส์ของตัวแปรปัจจัย 3 ตัว ได้แก่ time, cd4, age อย่างมีนัยสำคัญทางสถิติ 0.05 โดยค่าประมาณมาสร้างสมการลอจิสติกส์ (2.1-2.2) คือ $\text{logit } [P(Y \leq 1|x)] = -1.2536 - 0.00588\text{time}(1) + 0.3352\text{time}(2) - 0.2521\text{time}(3) + 0.3672\text{cd4} (1)$

$$- 0.9186\text{age}(1) + 0.3105\text{age}(2) + 0.6248\text{age}(3) \dots(2.1)$$

$$\text{ดังนั้น } \text{logit } [P(Y \leq 2|x)] = 0.8876 - 0.00588\text{time}(1) + 0.3352\text{time}(2) - 0.2521\text{time}(3) + 0.3672\text{cd4} (1) - 0.9186\text{age}(1) + 0.3105\text{age}(2) + 0.6248\text{age}(3) \dots(2.2)$$

ตารางที่ 8 ตารางการจำแนกผลการพยากรณ์ของตัวแบบลอจิสติกส์ 3 กลุ่ม (validation set)

ค่าสังเกต	ค่าพยากรณ์			
	ทำกิจกรรมได้ตามปกติ	เริ่มมีอาการแสดงของโรค	ทำกิจกรรมตามปกติไม่ได้	ร้อยละ
ทำกิจกรรมได้ตามปกติ	74	8	4	86.00
เริ่มมีอาการแสดงของโรค	16	82	5	80.00
ทำกิจกรรมตามปกติไม่ได้	10	11	21	50.00
ภาพรวม				76.60

จากตารางที่ 8 พบว่าความถูกต้องในการพยากรณ์ของตัวแบบลอจิสติกส์ที่มีตัวแปรตอบสนอง 3 กลุ่มแบบมีอันดับ (2.1-2.2) ในผู้ป่วยโรคเอดส์ที่ทำกิจกรรมได้ตามปกติจะพยากรณ์ได้ถูกต้องร้อยละ 86.00 ถ้าเป็นกลุ่มผู้ป่วยโรคเอดส์ที่ทำกิจกรรมได้ตามปกติแต่เริ่มมีอาการแสดงของโรคจะพยากรณ์ได้ถูกต้องร้อยละ 80.00 และถ้าเป็นกลุ่มผู้ป่วยโรคเอดส์ที่ไม่สามารถทำกิจกรรมได้ตามปกติจะพยากรณ์ได้ถูกต้องร้อยละ 50.00 ส่วนความถูกต้องในการพยากรณ์ในภาพรวม ตัวแบบนี้สามารถพยากรณ์ได้ถูกต้องร้อยละ 76.60 ซึ่งมีประสิทธิภาพใกล้เคียงกับข้อมูลที่ได้จาก training set ดังนั้นตัวแบบที่ได้นี้ จึงมีความน่าเชื่อถือในการนำไปใช้พยากรณ์กลุ่มต่อไป

4. สรุป อภิปรายผล และข้อเสนอแนะ

ในงานวิจัยนี้สนใจวิเคราะห์และศึกษาปัจจัยที่มีผลต่อการมีชีวิตอยู่และสภาวะของผู้ป่วยโรคเอดส์โดย

ใช้ตัวแบบลอจิสติกส์ 2 กลุ่ม และตัวแบบลอจิสติกส์ 3 กลุ่ม โดยที่ตัวแปรตอบสนอง แบ่งออกเป็น 2 กรณี ดังนี้

4.1 กรณีที่ 1 การวิเคราะห์ข้อมูลโดยใช้ตัวแบบลอจิสติกส์ 2 กลุ่ม (binary logit model)

การประเมินสถานการณ์การมีชีวิต แบ่งเป็น 2 กลุ่ม กลุ่มที่ 1 แทนมีชีวิตอยู่ กลุ่มที่ 2 แทนเสียชีวิต พบว่าตัวแบบเหมาะสมที่ได้จากวิธีการเลือกแบบขั้นตอนซึ่งมี 1 สมการ คือ $\text{logit } [P(Y=1|x)] = 4.0388 + 2.3187\text{dtime}(1) - 0.2178\text{dtime}(2) - 1.0313\text{dtime}(3) + 0.6671\text{tx}(1) + 0.6368\text{cd4}(1)$

พบว่าปัจจัยหรือตัวแปรอธิบายที่เลือกเข้าตัวแบบ ได้แก่ ระยะเวลาการมีชีวิตอยู่ของผู้ป่วย (dtime) ประเภทของการรักษาผู้ป่วย (tx) และระดับเซลล์เม็ดเลือดขาว (cd4) อย่างมีนัยสำคัญทางสถิติ 0.05 โดยตัวแบบมีภาวะสารรูปดี และประสิทธิภาพความในการพยากรณ์ได้ถูกต้องร้อยละ 98.70

4.2 กรณีที่ 2 การวิเคราะห์ข้อมูลโดยใช้ตัวแบบลอจิสติกสะสม (cumulative logit model)

การประเมินสภาวะผู้ป่วย แบ่งเป็น 3 กลุ่ม กลุ่มที่ 1 แทนผู้ป่วยสามารถทำกิจกรรมได้ตามปกติ กลุ่มที่ 2 แทนผู้ป่วยสามารถทำกิจกรรมได้ตามปกติแต่เริ่มมีอาการแสดงของโรค และกลุ่มที่ 3 แทนผู้ป่วยไม่สามารถทำกิจกรรมได้ตามปกติ พบว่าตัวแบบเหมาะสมที่ได้จากวิธีการเลือกแบบขั้นตอน 2 สมการ คือ $\text{logit}[P(Y \leq 1|x)] = -1.2536 - 0.00588\text{time}(1) + 0.3352\text{time}(2) - 0.2521\text{time}(3) + 0.3672\text{cd4}(1) - 0.9186\text{age}(1) + 0.3105\text{age}(2) + 0.6248\text{age}(3)$ และ $\text{logit}[P(Y \leq 2|x)] = 0.8876 - 0.00588\text{time}(1) + 0.3352\text{time}(2) - 0.2521\text{time}(3) + 0.3672\text{cd4}(1) - 0.9186\text{age}(1) + 0.3105\text{age}(2) + 0.6248\text{age}(3)$

พบว่าปัจจัยหรือตัวแปรอธิบายที่เลือกเข้าตัวแบบ ได้แก่ ระยะเวลาตั้งแต่วินิจฉัยการเป็นโรค เอ็ดส์จนกระทั่งเสียชีวิตหรือสิ้นสุดการทดลอง (time) ระดับเซลล์เม็ดเลือดขาว (cd4) และอายุ (age) อย่างมีนัยสำคัญทางสถิติ 0.05 โดยตัวแบบมีภาวะสารูปดีและมีประสิทธิภาพความในการพยากรณ์ได้ถูกต้องร้อยละ 76.60

4.3 อภิปรายผลและข้อเสนอแนะ

เนื่องจากการวิเคราะห์การมีชีวิตรอยู่และสภาวะกิจกรรมของผู้ป่วยโรคเอดส์ที่พบความเกี่ยวพันระหว่างปัจจัยจากตัวแบบลอจิสติกและตัวแบบลอจิสติกสะสมนั้น อาจให้ผลลัพธ์ที่แตกต่างกันได้ เนื่องจากการใช้ตัวแปรตอบสนองคนละแบบกันขึ้นอยู่กับวัตถุประสงค์ของการนำไปใช้สำหรับตัวแปรตอบสนองแบบใดด้วย ดังนั้นสำหรับปัจจัยที่ส่งผลต่อความน่าจะเป็นของการมีชีวิตรอยู่และการประเมินสภาวะของผู้ป่วยโรคเอดส์ระดับต่าง ๆ จึงมีความเป็นไปได้ว่าอาจให้ความสำคัญของปัจจัยต่าง ๆ จากทั้ง 2 ตัวแบบที่ส่งผลต่อความน่าจะเป็นให้ผู้ป่วยโรคเอดส์สามารถมีชีวิตรอยู่

ยาวต่อไป หรือมีสภาวะกิจกรรมของผู้ป่วยโรคเอดส์แบบต่าง ๆ อย่างมีนัยสำคัญทางสถิติ [10] ส่วนการวิจัยครั้งต่อไปอาจสร้างตัวแบบทางเลือกอื่นที่อยู่ในกลุ่มของตัวแบบเชิงเส้นน้อยตัวไป และอาจใช้ปัจจัยเสี่ยงเดิมและปัจจัยที่เกี่ยวข้องในด้านอื่นเพิ่มเติม เช่น พฤติกรรมในการใช้ชีวิตประจำวันของผู้ป่วย ความรับผิดชอบในการดูแลตนเอง วิธีการรักษา [11] ดังนั้นหากผู้สนใจการวิเคราะห์ด้วยตัวแบบดังกล่าวเพื่อนำมาประยุกต์ให้เหมาะสมกับผู้ป่วยโรคเอดส์ในประเทศไทย อาจใช้ข้อมูลให้หลากหลาย [11,12] โดยให้มีขนาดตัวอย่างที่มากพอ รวมทั้งสามารถทดลองกับตัวแบบ GLMs ทางเลือกอื่น ๆ ที่เหมาะสมด้วย

5. รายการอ้างอิง

- [1] ประพันธ์ ภาณุภาค และคณะ, 2553, แนวทางการตรวจวินิจฉัยและการดูแลรักษาผู้ติดเชื้อเอชไอวีและผู้ป่วยเอดส์ระดับชาติ ปี พ.ศ. 2553, พิมพ์ครั้งที่ 1, โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย จำกัด, 480 น.
- [2] เชิดเกียรติ แก้วกลสิกิจ, 2550, ความรู้เบื้องต้นเกี่ยวกับโรคเอดส์, แหล่งที่มา : <http://dpc9.ddc.moph.go.th/napha9/aidsdata/aidsbas ic.html>, 4 มกราคม 2555.
- [3] Hammer, S.M., Squires, K.E., Hughes, M.D., Grimes, J.M., Demeter, L.M., Currier, J.S., Eron, J.J. Jr, Feinberg, J.E, Balfour, H.H. Jr, Deyton, L.R., Chodakewitz, J.A., Fischl, M.A., Phair, J.P., Pedneault, L., Nguyen, B., and Cook, J.C., 1997, A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection

- and CD4 cell counts of 200 per cubic millimeter or less, AIDS Clinical Trials Group 320 Study Team, N Engl. J. Med. 337: 725-733.
- [4] Agresti, A., 2007, Introduction to Categorical Data Analysis, John Wiley & Sons, Inc., New York, 372 p.
- [5] Agresti, A., 2013, Categorical Data Analysis, John Wiley & Sons, Inc., New York, 714 p.
- [6] McCullagh, P., 1980, Regression models for ordinal data, J. Royal Statist. B42: 109-142.
- [7] Pongsapukdee, V. and Sukgumphaphan, S., 2008, Three cumulative link models for ordinal responses: A review of methods and power comparisons, Songklanakarin J. Sci. Technol. 30: 805-811.
- [8] SAS (Enterprise Guide 5.1), 2013, Statistical Analysis System Software, Enterprise Guide 5.1 License of Department of statistics, Faculty of Science, Silpakorn University.
- [9] วีรานันท์ พงศาภักดี, 2555, การวิเคราะห์ข้อมูลจำแนกประเภท : ทฤษฎีและการประยุกต์ด้วย GLIM, SPSS, SAS และ MTB, พิมพ์ครั้งที่ 3, โรงพิมพ์มหาวิทยาลัยศิลปากร, นครปฐม, 495 น.
- [10] ชัยญานุช อักษร, เพชรี อ่อนลออ และวีรานันท์ พงศาภักดี, 2557, การพยากรณ์ความน่าจะเป็นการไม่สำเร็จการศึกษาภายใน 4 ปี ภายใต้ตัวแบบลอจิต 2 กลุ่ม และตัวแบบลอจิตออกดส์เชิงสัดส่วน, ว.วิทยาศาสตร์และเทคโนโลยี 22: 814-827.
- [11] คันธพจน์ สมนึกชัยเจริญ, 2552, ใจใกล้ใจ โยงใยถึงปัญหา พัฒนาการเก็บข้อมูลใหม่ก้าวอย่างอย่างไร เข้าใจผู้ป่วยโรคเอดส์, แหล่งที่มา : <http://www.photharam.com/ptrhos/attachments/article/30/pdf16.pdf>, 12 กุมภาพันธ์ 2557.
- [12] พิพิธภัณฑ์สภากาชาดไทย, 2555, 3 ทศวรรษโรคเอดส์ อดีต ปัจจุบัน และอนาคต, แหล่งที่มา : <http://thairedcross.livingmuseum.sc.chula.ac.th/?p=340>, 10 กุมภาพันธ์ 2557.