

การศึกษาปัจจัยที่ส่งผลต่อระดับพลาสมาเบต้าแคโรทีน
และระดับพลาสมาเรตินอลด้วยตัวแบบโลจิสติก
ปานจิต วัฒนสารัช เบญจวรรณ ช่างฟัง และ วีรานันท์ พงศาภักดิ์*

**The Study of Factors Affecting the Levels of Plasma Beta-Carotene
and Plasma Retinol Using the Logistic Model**

Parnchit Wattanasaruch, Benjawan Changpueng and Veeranun Pongsapakdee*

ภาคสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร
วิทยาเขตพระราชวังสนามจันทร์ จังหวัดนครปฐม 73000

*Corresponding author E-mail: veeranun@su.ac.th

บทคัดย่อ

การศึกษาปัจจัยที่เกี่ยวข้องกับระดับพลาสมาเบต้าแคโรทีนและระดับพลาสมาเรตินอลด้วยตัวแบบโลจิสติกนี้เป็นการวิเคราะห์กับข้อมูลจริงโดยใช้ข้อมูลจากเว็บไซต์ <http://lib.stat.cmu.edu/> ซึ่งเป็นคนไข้ที่ได้รับการสนับสนุนเบื้องต้นว่าอาจเป็นมะเร็ง เพื่อศึกษาอิทธิพลของตัวแปรอธิบายต่าง ๆ ที่มีผลต่อตัวแปรตอบสนอง (Y) คือพลาสมาเบต้าแคโรทีนและพลาสมาเรตินอล ซึ่งมีการศึกษาทางการแพทย์พบว่าเบต้าแคโรทีนจะเปลี่ยนแปลงในกระแสเลือดกลายเป็นวิตามินเอหรือพลาสมาเรตินอล ใช้เป็นสารต้านอนุมูลอิสระ (Antioxidants) และอาจเป็นสาเหตุหนึ่งของการเกิดโรคมะเร็ง ตัวแปรตอบสนอง (Y) แบ่งเป็นสองกลุ่มคือกลุ่มของคนไข้ที่มีระดับพลาสมาเบต้าแคโรทีนน้อยกว่า 243 ng/ml กับกลุ่มของคนไข้ที่มีระดับพลาสมาเบต้าแคโรทีนมากกว่าหรือเท่ากับ 243 ng/ml และกลุ่มของคนไข้ที่มีระดับพลาสมาเรตินอลน้อยกว่า 555 ng/ml กับกลุ่มของคนไข้ที่มีระดับพลาสมาเรตินอลมากกว่าหรือเท่ากับ 555 ng/ml ส่วนตัวแปรอธิบายได้แก่อายุ (Age) เพศ (Sex) การสูบบุหรี่ (Smoke) ดัชนีความหนาของร่างกาย (Quetelet) การบริโภควิตามินเสริม (Vituse) ปริมาณแคลอรีที่บริโภคต่อวัน (Calories) ปริมาณไขมันที่บริโภคต่อวัน (Fat) ปริมาณใยอาหารที่บริโภคต่อวัน (Fiber) จำนวนครั้งของการดื่มแอลกอฮอล์ต่อสัปดาห์ (Alcohol) ปริมาณคอเลสเตอรอลที่บริโภคต่อวัน (Cholesterol) เบต้าแคโรทีนที่บริโภคต่อวัน (Betadiet) เรตินอลที่บริโภคต่อวัน (Retinoldiet) โดยใช้

วิธีการคัดเลือกตัวแปรแบบขั้นตอน (Stepwise) และประมวลผลด้วยโปรแกรม SAS (Statistical Algorithm Systems Version 9.0)

ผลการวิจัยพบว่าตัวแปรอธิบายที่ส่งผลต่อระดับพลาสมาเบต้าแคโรทีนคือ อายุ (*Age*) เบต้าแคโรทีนที่บริโภคต่อวัน (*Betadiet*) คำนีความหนาของร่างกาย (*Quetelet*) เรตินอลที่บริโภคต่อวัน (*Retinoldiet*) การสูบบุหรี่ (*Smoke*) การบริโภควิตามิน (*Vituse*) และตัวแบบโลจิสติกที่ได้คือ

$$\begin{aligned} \logit \hat{P}(x) = & 0.8072 - 0.0397 \text{ Age} - 0.0004 \text{ Betadiet} \\ & + 0.1435 \text{ Quetelet} + 0.0006 \text{ Retinoldiet} \\ & - 0.6482 \text{ Smoke (1)} - 0.3744 \text{ Smoke (2)} \\ & - 0.5300 \text{ Vituse (1)} - 0.2575 \text{ Vituse (2)} \end{aligned}$$

ตัวแปรอธิบายที่ส่งผลต่อระดับพลาสมาเรตินอล คือ อายุ (*Age*) จำนวนครั้งของการดื่มแอลกอฮอล์ต่อสัปดาห์ (*Alcohol*) ปริมาณใยอาหารที่บริโภคต่อวัน (*Fiber*) คำนีความหนาของร่างกาย (*Quetelet*) และตัวแบบโลจิสติกที่ประมาณได้คือ

$$\logit \hat{P}(x) = 2.1183 - 0.0332 \text{ Age} - 0.0887 \text{ Alcohol} + 0.0252 \text{ Fiber} - 0.0285 \text{ Quetelet}$$

คำสำคัญ: ความน่าจะเป็นแบบมีเงื่อนไข พลาสมาเบต้าแคโรทีน พลาสมาเรตินอล ตัวแบบโลจิสติก

Abstract

The study of plasma beta-carotene and plasma retinol levels data obtained from website <http://lib.stat.cmu.edu/> on "Determinants of Plasma Retinol and Beta-Carotene Levels", department of Statistics, Carnegie Mellon University (2001) Pittsburgh, PA 15213. The experimental units or patients, who had consulted at the beginning, might take the risk of being cancer. The research objective was to study the explanatory variables affecting the response variables (Y), each of which was the plasma beta-carotene and plasma retinol levels. There were medical evidences that beta-carotene will be modified to be vitamin A or plasma retinol in blood. They were antioxidants which may cause cancer. The response variable (Y) was divided into two groups. One was patients who had the plasma beta-carotene level less than 243 ng/ml, other of which the plasma beta-carotene level more than and equal to 243 ng/ml. As well as the plasma retinol level less than 555 ng/ml represented the first group and the other group was the plasma retinol level more than and equal to 555 ng/ml. The explanatory variables were *Age*, *Sex*, Smoking status (*Smoke*), *Quetelet*, vitamin use

(*Vituse*), Number of calories consumed/day (*Calories*), Grams of fat consumed/day (*Fat*), Fiber consumed/day (*Fiber*), Number of alcoholic drinks consumed/week (*Alcohol*), Cholesterol consumed/day(*Cholesterol*), Dietary beta-carotene consumed (*Betadiet*) and Dietary retinol consumed (*Retinoidiet*). The data analysis was based on logistic model. The selection of models under stepwise selection method was performed using SAS version 9.0

The results revealed that the statistically significant factors affecting plasma beta-carotene levels were *Age*, Dietary beta-carotene (*Betadiet*), *Quetelet*, Dietary retinol consumed (*Retinoidiet*), Smoking status (*Smoke*), and Vitamin use (*Vituse*). The estimated logit model was

$$\begin{aligned} \log it \hat{P}(x) = & 0.8072 - 0.0397Age - 0.0004Betadiet \\ & + 0.1435 Quetelet + 0.0006 Re tinoldiet \\ & - 0.6482 Smoke (1) - 0.3744 Smoke (2) \\ & - 0.5300Vituse (1) - 0.2575 Vituse (2). \end{aligned}$$

The affecting factors to plasma retinol levels were *Age*, Number of alcoholic drinks consumed/week (*Alcohol*), Fiber consumed/day (*Fiber*) and *Quetelet*. The estimated logit model was

$$\begin{aligned} \log it \hat{P}(x) = & 2.1183 - 0.0332Age - 0.0887 Alcohol \\ & + 0.0252 Fiber - 0.0285 Quetelet . \end{aligned}$$

Keywords: Conditional Probability Plasma beta-carotene Plasma Retinol Logistic Model

บทนำ

ปัจจุบันการรักษาโรคต่างๆ เป็นปัญหาทางสาธารณสุขที่สำคัญอย่างหนึ่งของประเทศ โดยเฉพาะปัญหาเกี่ยวกับโรคมะเร็ง เนื่องจากมะเร็งเป็นหนึ่งในสาเหตุสำคัญของการเสียชีวิตในคนทั่วโลก สาเหตุของการเกิดโรคมะเร็งเกิดจากสิ่งแวดล้อมหรือภายนอกในร่างกายเช่น การรับประทานอาหารและเครื่องดื่มที่มีสารก่อมะเร็งปนเปื้อนในอาหารหรือจากพฤติกรรมบางอย่างเช่น การสูบบุหรี่ การดื่มแอลกอฮอล์ หรือเกิดจากความผิดปกติภายในร่างกายเช่น การมีภูมิคุ้มกันบกพร่องหรือภาวะทุพโภชนาการ ส่วนวิธีการป้องกันเพื่อลดความเสี่ยงต่อการเป็นโรคมะเร็งคือ เลือกรับประทานอาหารที่มีสารช่วยต่อต้านอนุมูลอิสระเช่น วิตามินเอ เบต้าแคโรทีน วิตามินซี งดการสูบบุหรี่และดื่มแอลกอฮอล์

เบต้าแคโรทีน (Beta-carotene) เป็นสารตั้งต้นของวิตามินเอ (โปรวิตามินเอ) เมื่อเบต้าแคโรทีนเข้าสู่ร่างกายจะเปลี่ยนแปลงในกระแสเลือดกลายเป็นวิตามินเอ (Retinol) หรือพลาสมาเรตินอลมีบทบาทสำคัญคือ เป็นสารต้านอนุมูลอิสระ (Antioxidants) ซึ่งเป็นสาเหตุของการเกิดโรคมะเร็ง หาก

ร่างกายมีระดับพลาสมาเบต้าแคโรทีนและพลาสมาเรตินอลต่ำจะส่งผลต่อระบบภูมิคุ้มกันของร่างกาย (สรจักร ศิริบริรักษ์, 2007)

ตัวแบบโลจิสติกคือตัวแบบที่ใช้แสดงความสัมพันธ์ระหว่างตัวแปรอธิบายกับความน่าจะเป็นแบบมีเงื่อนไขของตัวแปรตอบสนอง (วีรานันท์ พงศาภักดิ์, 2541) ใช้ในการอธิบายและพยากรณ์ความน่าจะเป็นแบบมีเงื่อนไขของตัวแปรตอบสนอง Y หรือ $P(Y = 1 | x) = P(x)$ เมื่อตัวแปรตอบสนองเป็นตัวแปรเชิงกลุ่ม (Categorical) โดยทั่วไปเป็นแบบเชิงกลุ่ม Nominal และ Ordinal ส่วนตัวแปรอธิบายอาจเป็นตัวแปรเชิงกลุ่มหรือตัวแปรต่อเนื่องก็ได้

การวิเคราะห์ตัวแบบโลจิสติกจะทำการแปลงค่าของ $P(x)$ หรือค่าคาดหวังแบบมีเงื่อนไขของ Y เมื่อกำหนดค่าของ X ซึ่งกำหนดให้ $P(x) = E(Y | x)$ คือค่าเฉลี่ยแบบมีเงื่อนไขของตัวแปรตอบสนอง Y ซึ่งเป็นตัวแปรเชิงกลุ่ม เช่น 2 กลุ่ม (Binary) หรือหลายกลุ่ม (Polychotomous) เมื่อกำหนดตัวแปรอธิบาย X มาให้ (Conditional mean of given x) การแจกแจงของตัวแปรตอบสนอง (Y) ที่มี 2 กลุ่ม เป็นแบบแบร์นูลลี (Bernoulli distribution) หรือ $Ber(P(x))$ เมื่อพิจารณาที่ยังไม่ได้จัดกลุ่ม (Ungrouped data) และเป็น $Bin(n, P(x))$ เมื่อพิจารณาข้อมูลแบบจัดกลุ่ม (Grouped data) โดย $P(x)$ แทนความน่าจะเป็นของตัวแปรตอบสนองในกลุ่มที่สนใจศึกษา ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรเชิงกลุ่ม 2 ระดับ คือ $Y = 0$ เป็นเหตุการณ์ที่ไม่สนใจและ $Y = 1$ เป็นเหตุการณ์ที่สนใจ เมื่อ $Y = 1$ ความน่าจะเป็นของตัวแปรตอบสนองคือ $P(x)$ และเมื่อ $Y = 0$ ความน่าจะเป็นคือ $1 - P(x)$

ตัวแบบถดถอยโลจิสติกคือ

$$P(x) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}}$$

และเพื่อให้ค่าของ $P(x)$ หรือ $E(Y | X)$ ที่มีค่าอยู่ระหว่าง 0 ถึง 1 และค่าของ x สามารถมีได้ไม่จำกัด จึงมีความจำเป็นต้องแปลง $P(x)$ ให้อยู่ในรูปอื่นที่เข้าใจง่ายและมีคุณสมบัติตามต้องการ ซึ่งมีรูปแบบดังนี้

$$Q(x) = \log \left[\frac{P(x)}{1 - P(x)} \right] = \beta_0 + \beta X$$

โดยที่ คือ ฟังก์ชันเชิงเส้นในเทอมของพารามิเตอร์ซึ่งมีค่าต่อเนื่องหรืออาจมีค่าในช่วง $-\infty$ ถึง $+\infty$ โดยขึ้นอยู่กับค่าของ X

$$\left[\frac{P(x)}{1 - P(x)} \right]$$

คือ อัตราส่วน Odds ซึ่งหมายถึง อัตราส่วนระหว่างความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจกับความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ที่สนใจ

$$\log \left[\frac{P(x)}{1-P(x)} \right] \text{ คือ โลจิทของ } P(x) \text{ หรือ } \text{logit}P(x)$$

รูปแบบโลจิทของ $P(x)$ สามารถขยายนำไปใช้สำหรับ Y ที่มีมากกว่า 2 กลุ่ม และใช้สำหรับ X ที่มีหลายตัวแปรทั้งแบบเชิงกลุ่มและไม่เชิงกลุ่ม โดยเรียกความสัมพันธ์ระหว่างโลจิทของ $P(x)$ กับผลรวมเชิงเส้นของ X 's ว่า ตัวแบบโลจิสติก (Logistic model) ซึ่งเป็นฟังก์ชันเชิงเส้นแสดงความสัมพันธ์ระหว่างตัวแปรอธิบายกับตัวแปรตามที่งานวิจัยนี้พิจารณาให้มีค่าเพียงสองค่าคือ 1 กับ 0

วัตถุประสงค์ของการวิจัย

1. พยากรณ์ความน่าจะเป็นที่คนไข้จะเสี่ยงต่อระดับพลาสมาเบต้าแคโรทีนและระดับพลาสมาเรตินอลต่ำ
2. ทราบปัจจัยที่ส่งผลต่อระดับพลาสมาเบต้าแคโรทีนและระดับพลาสมาเรตินอล

ขอบเขตของการวิจัย

งานวิจัยนี้ผู้วิจัยสนใจศึกษาการวิเคราะห์ข้อมูลด้วยตัวแบบโลจิสติก เมื่อกำหนดให้ตัวแปรตอบสนองแบ่งออกเป็น 2 ระดับ (Binary) โดยใช้ข้อมูลจริงจากเว็บไซต์ <http://lib.stat.cmu.edu/> ซึ่งเป็นการเก็บตัวอย่างของคนไข้จำนวน 313 คน

1. ตัวแปรตอบสนองที่ใช้ในงานวิจัยนี้คือ ตัวแปรเชิงกลุ่ม 2 กลุ่ม
 - 1.1 ระดับของพลาสมาเบต้าแคโรทีนที่ต่ำกว่าและสูงกว่าหรือเท่ากับ 243 ng/ml ที่วัดได้จากเลือดโดยกำหนดให้เป็นกลุ่ม 1 และ 0 ตามลำดับ
 - 1.2 ระดับพลาสมาเรตินอลที่ต่ำกว่าและสูงกว่าหรือเท่ากับ 555 ng/ml ที่วัดได้จากเลือดโดยกำหนดให้เป็นกลุ่ม 1 และ 0 ตามลำดับ
2. เกณฑ์การแบ่งกลุ่มในข้อ 1. จากงานวิจัยเรื่อง Diurnal variation in plasma levels of retinol, tocopherol, and Beta-carotene ซึ่งวัดระดับพลาสมาเบต้าแคโรทีนและระดับพลาสมาเรตินอล จากคนปกติที่มีสุขภาพดี 15 คน โดย Nierenberg DW และ Stukel (1987)
3. ตัวแปรอธิบายที่ใช้ในงานวิจัยนี้มีทั้งหมด 12 ตัว ได้แก่
 - 3.1 อายุ (Age)
 - 3.2 เพศ (Sex)
 - 3.3 การสูบบุหรี่ (Smoke)
 - ไม่เคยสูบบุหรี่ = 1
 - เคยสูบบุหรี่แต่ปัจจุบันเลิกแล้ว = 2
 - สูบบุหรี่อย่างสม่ำเสมอ = 3

3.4 ดัชนีความหนาของร่างกาย (*Quetelet*)

3.5 การบริโภควิตามิน (*Vituse*)

- บริโภควิตามินเสริมเป็นประจำ = 1
- บริโภควิตามินเสริมไม่ประจำ = 2
- ไม่เคยบริโภควิตามินเสริม = 3

3.6 ปริมาณแคลอรีที่บริโภคต่อวัน (*Calories*)

3.7 ปริมาณไขมันที่บริโภคต่อวันมีหน่วยเป็นกรัม (*Fat*)

3.8 ปริมาณใยอาหารที่บริโภคต่อวัน (*Fiber*)

3.9 จำนวนครั้งของการดื่มแอลกอฮอล์ต่อสัปดาห์ (*Alcoholic*)

3.10 ปริมาณคอเลสเตอรอลที่บริโภคต่อวันมีหน่วยเป็นมิลลิกรัม (*Cholesterol*)

3.11 เบต้าแคโรทีนที่บริโภคต่อวัน (*Betadiet*) มีหน่วยเป็นไมโครกรัมต่อวัน

3.12 เรตินอลที่บริโภคต่อวัน (*Retinoldiet*) มีหน่วยเป็นไมโครกรัมต่อวัน

การพัฒนาตัวแบบ

1. ทำการคัดเลือกตัวแปรอธิบายเพื่อเข้าในตัวแบบโลจิสติก โดยใช้วิธีการเลือกตัวแปรแบบขั้นตอนโดยวิธีอัตราส่วนภาวะน่าจะเป็น (Stepwise (likelihood ratio : LR))

2. จากตัวแปรอธิบายที่คัดเลือกได้ในข้อ 1 จะนำตัวแปรดังกล่าวมาทำการประมาณค่าพารามิเตอร์ของตัวแบบ โลจิสติกด้วยวิธีภาวะน่าจะเป็นสูงสุด(Maximum Likelihood or ML)

3. นำค่าประมาณพารามิเตอร์ของตัวแบบ โลจิสติกที่ได้จากข้อ 2 มาสร้างตัวแบบ โลจิสติก

4. จากนั้นทำการวัดความสัมพันธ์และการวัดภาวะสารูปดีของตัวแบบด้วย Percent Concordant, Percent Discordant, c, Akaike' Information Criterion (AIC) และ Bayesian Information Criterion (BIC)

5. นำตัวแบบ โลจิสติกที่ได้มาใช้ในการพยากรณ์ความน่าจะเป็นที่คนไข้จะเสี่ยงต่อการมีระดับพลาสมาเบต้าแคโรทีนต่ำและความน่าจะเป็นที่คนไข้จะเสี่ยงต่อการมีระดับพลาสมาเรตินอลต่ำ

6. ตรวจสอบภาวะสารูปดีของตัวแบบและการวิเคราะห์ส่วนเหลือ

7. สรุปผลลัพธ์ที่ได้จากการวิเคราะห์

ผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลแบ่งเป็น 2 ตอนดังนี้

ตอนที่ 1 การวิเคราะห์ข้อมูลด้วยตัวแบบ โลจิสติก เมื่อกำหนดให้ตัวแปรตอบสนองคือระดับพลาสมาเบต้าแคโรทีน

ตอนที่ 2 การวิเคราะห์ข้อมูลด้วยตัวแบบโลจิสติก เมื่อกำหนดให้ตัวแปรตอบสนองคือระดับพลาสมาเรตินอล

ตาราง 1 ค่าประมาณพารามิเตอร์ของตัวแบบโลจิสติก ความคลาดเคลื่อนมาตรฐาน กรณีตัวแปรตอบสนองคือระดับพลาสมาเบต้าแคโรทีน

Parameter	Estimate	Standard Error	Wald Chi-Square	DF
Intercept	0.8072	1.0808	0.5578	1
Age	-0.0397	0.0113	12.3984	1
Betadiet	-0.0004	0.0001	17.7768	1
Quetelet	0.1435	0.0382	14.0921	1
Retinoldiet	0.0006	0.0004	2.5945	1
Smoke(1)	-0.6482	0.3029	4.5801	1
Smoke(2)	-0.3744	0.3143	1.4185	1
Vituse(1)	-0.5300	0.2132	6.1775	1
Vituse(2)	-0.2575	0.2508	1.0542	1

ตัวแบบที่ประมาณได้ (1) คือ

$$\begin{aligned} \log it \hat{P}(x) = & 0.8072 - 0.0397 Age - 0.0004 Betadiet \\ & + 0.1435 Quetelet + 0.0006 Retinoldiet \\ & - 0.6482 Smoke(1) - 0.3744 Smoke(2) \\ & - 0.5300 Vituse(1) - 0.2575 Vituse(2) \dots\dots\dots(1) \end{aligned}$$

ตาราง 2 ค่าประมาณ Odds Ratio มาตรฐาน กรณีตัวแปรตอบสนองคือระดับพลาสมาเบต้าแคโรทีน

Effect	Point Estimate	95% Wald Confidence Limits
Age	0.961	0.940-0.983
Betadiet	1.000	0.999-1.000
Quetelet	1.154	1.071-1.244
Retinoldiet	1.001	1.000-1.001
Smoke 1VS 3	0.188	0.040-0.888
Smoke 2 VS 3	0.247	0.051-1.199
Vituse 1VS 3	0.268	0.121-0.593
Vituse 2VS 3	0.352	0.141-0.876

การตีความหมายของตัวแปรอธิบายในตัวแบบที่ประมาณได้ (1) สำหรับระดับพลาสมาเบต้าแคโรทีน
กรณีตัวแปรอธิบายเป็นตัวแปรแบบต่อเนื่อง

- ตัวแปรอายุ (Age): คนไขที่มีอายุเพิ่มขึ้น 1 หน่วย ส่งผลให้ความเสี่ยงต่อระดับพลาสมาเบต้าแคโรทีนต่ำ เป็น 0.961 เท่าของความเสี่ยงต่อระดับพลาสมาเบต้าแคโรทีนต่ำของคนไขที่มีอายุเท่าเดิม เมื่อตัวแปรอธิบายตัวอื่นคงเดิม

- ตัวแปรดัชนีความหนาของร่างกาย (Quetelet): คนไขที่มีดัชนีความหนาของร่างกายน้อยลง 1 หน่วยส่งผลให้ความเสี่ยงต่อระดับพลาสมาเบต้าแคโรทีนต่ำ เป็น 1.154 เท่าของความเสี่ยงต่อระดับพลาสมาเบต้าแคโรทีนต่ำของคนไขที่มีดัชนีความหนาของร่างกายเท่าเดิม เมื่อตัวแปรอธิบายตัวอื่นคงเดิม

กรณีตัวแปรอธิบายเป็นตัวแปรแบบเชิงกลุ่ม

- คนไข้ที่ไม่เคยสูบบุหรี่ (1) เลขจะมีความเสี่ยงต่อการมีระดับพลาสมาเบต้าแคโรทีนต่ำ เป็น 0.188 เท่าเมื่อเทียบกับคนไข้ที่สูบบุหรี่อย่างสม่ำเสมอ (3) และค่าประมาณ Odds Ratio มีค่าน้อยกว่า 1 หมายความว่าคนไข้ที่สูบบุหรี่อย่างสม่ำเสมอ (3) มีความเสี่ยงต่อการมีระดับพลาสมาเบต้าแคโรทีนต่ำมากกว่าคนไข้ที่ไม่เคยสูบบุหรี่เลย (1)

- คนไข้ที่เคยสูบบุหรี่แต่ปัจจุบันเลิกแล้ว (2) จะมีความเสี่ยงต่อการมีระดับพลาสมาเบต้าแคโรทีนต่ำเป็น 0.247 เท่าของคนไข้ที่สูบบุหรี่อย่างสม่ำเสมอ (3) และค่าประมาณ Odds Ratio มีค่าน้อยกว่า 1 หมายความว่าคนไข้ที่สูบบุหรี่อย่างสม่ำเสมอ (3) มีความเสี่ยงต่อการมีระดับพลาสมาเบต้าแคโรทีนต่ำมากกว่าคนไข้ที่เคยสูบบุหรี่แต่ปัจจุบันเลิกแล้ว (2)

ตาราง 3 การทดสอบสมมติฐาน $H_0 : (\beta_i \neq 0)$ อย่างน้อย 1 ค่าของตัวแบบโลจิสติกด้วยตัวสถิติ Likelihood Ratio, Score และ Wald กรณีตัวแปรตอบสนองคือระดับพลาสมาเบต้าแคโรทีน

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	80.1873	8	<0.0001*
Score	69.7134	8	<0.0001*
Wald	51.7133	8	<0.0001*

*นัยสำคัญทางสถิติที่ระดับ 0.05

จากตาราง 3 พบว่าค่าสถิติ Likelihood Ratio หรือค่า $\Delta Deviance$ เท่ากับ 80.1873, ค่าสถิติ Score เท่ากับ 69.7134 และค่าสถิติ Wald เท่ากับ 51.7133 ต่างมี P-values < 0.0001 หมายความว่าค่าประมาณของพารามิเตอร์ในตัวแบบโลจิสติกมีค่าไม่เท่ากับศูนย์ ภายใต้ระดับนัยสำคัญทางสถิติ 0.05 (P-value<0.0001) กล่าวอีกนัยหนึ่งคือ ตัวแปรที่ได้จากการคัดเลือกตัวแปรแบบขั้นตอน

(Stepwise) ซึ่งประกอบไปด้วยปัจจัย Age, Betadiet , Quetelet , Smoke , Vituse และ Retinoldiet (ตาราง 1) นั้นมีความเหมาะสมที่จะใช้ในการอธิบายระดับพลาสมาเบต้าแคโรทีนและการพยากรณ์

ตาราง 4 การวัดความสัมพันธ์ของตัวแบบ

ตัวสถิติ	ค่าของความสัมพันธ์
Percent Concordant	81.80
Percent Discordant	17.90
c	0.820

ตาราง 5 ประสิทธิภาพในการพยากรณ์ของตัวแบบ

Observed	Predicted			
	Plasma Beta-carotene Levels		Percentage	
	1	0	Correct	
Plasma Beta-carotene Levels	1	231	44	94.7
	0	13	25	36.2
Overall Percentage				81.8

ตาราง 6 การวัดภาวะสารูปคีของตัวแบบ

Criterion	Intercept and Covariates
AIC	268.011
BIC	290.234

ระดับพลาสมาเบต้าแคโรทีน พบว่าตัวแบบยังให้ค่าร้อยละของความสอดคล้องก่อนข้างสูง (81.80) และค่าแสดงความสัมพันธ์ของตัวแปรในตัวแบบได้สูงด้วย (ตาราง 4) และจากตัวแบบที่ประมาณได้ (1) สามารถพยากรณ์ในภาพรวมได้ถูกต้อง 81.8% (ตาราง 5) นอกจากนี้การตรวจสอบค่าส่วนเหลือ (Residuals deviance) ของตัวแบบพบว่าส่วนเหลือมีค่าอยู่ระหว่าง [-2,2] และมีการกระจายตัวค่อนข้างดี หมายความว่า ตัวแบบมีความเหมาะสมกับข้อมูล

ตาราง 7 ค่าประมาณพารามิเตอร์ของตัวแบบโลจิสติก ความคลาดเคลื่อนมาตรฐาน กรณีตัวแปรตอบสนองคือระดับพลาสมาเรตินอล

Parameter	Estimate	Standard Error	Wald Chi-Square	DF
Intercept	2.1183	0.7632	7.7030	1
Age	-0.0332	0.0084	15.6878	1
Alcohol	-0.0887	0.0311	8.1517	1
Fiber	0.0252	0.0222	1.2897	1
Quetelet	-0.0285	0.0200	2.0336	1

ตัวแบบที่ประมาณได้ (2) คือ

$$\logit \hat{P}(x) = 2.1183 - 0.0332Age - 0.0887Alcohol + 0.0252Fiber - 0.0285Quetelet \dots\dots\dots(2)$$

ตาราง 8 ค่าประมาณ Odds Ratio กรณีตัวแปรตอบสนองคือระดับพลาสมาเรตินอล

Effect	Point Estimate	95% Wald Confidence Limits
Age	0.9670	0.9520-0.9830
Alcohol	0.9150	0.8610-0.9730
Fiber	1.0260	0.9340-1.0110
Quetelet	0.9720	0.9340-1.0110

การตีความหมายของตัวแปรอธิบายในตัวแบบที่ประมาณได้ (2) สำหรับระดับพลาสมาเรตินอล

- ตัวแปรปริมาณใยอาหารที่บริโภคต่อวัน (*Fiber*): คนไข้ที่มีปริมาณใยอาหารที่บริโภคต่อวันเพิ่มขึ้น 1 หน่วย ส่งผลให้ความเสี่ยงต่อระดับพลาสมาเรตินอลต่ำ เป็น 1.026 เท่าของความเสี่ยงต่อระดับพลาสมาเรตินอลต่ำของคนไข้ที่มีปริมาณใยอาหารที่บริโภคต่อวันเท่าเดิม เมื่อตัวแปรอธิบายตัวอื่นคงเดิม
- ตัวแปรดัชนีความหนาของร่างกาย (*Quetelet*): คนไข้ที่มีดัชนีความหนาของร่างกายเพิ่มขึ้น 1 หน่วย ส่งผลให้ความเสี่ยงต่อระดับพลาสมาเรตินอลต่ำ เป็น 0.972 เท่าของความเสี่ยงต่อระดับพลาสมาเรตินอลต่ำของคนไข้ที่มีดัชนีความหนาของร่างกายเท่าเดิม เมื่อตัวแปรอธิบายตัวอื่นคงเดิม

ตาราง 9 การทดสอบสมมติฐาน $H_0 : (\beta_i \neq 0)$ อย่างน้อย 1 ค่าของตัวแบบโลจิสติก ด้วยตัวสถิติ Likelihood Ratio, Score และ Wald กรณีตัวแปรตอบสนองคือระดับพลาสมาเรตินอล

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	28.5042	4	<0.0001
Score	26.6108	4	<0.0001
Wald	24.2785	4	<0.0001

*นัยสำคัญทางสถิติที่ระดับ 0.05

จากตาราง 9 พบว่าค่าสถิติ Likelihood Ratio หรือค่า $\Delta Deviance$ เท่ากับ 28.042 ค่าสถิติ Score เท่ากับ 26.619 และค่าสถิติ Wald เท่ากับ 24.2785 หมายความว่าค่าประมาณของพารามิเตอร์ในตัวแบบโลจิสติกมีนัยสำคัญและมีค่าไม่เท่ากับศูนย์ ภายใต้ระดับนัยสำคัญ 0.05 (P-value<0.0001) ดังนั้น ตัวแปรที่ใช้ระดับพลาสมาเรตินอล ประกอบไปด้วยปัจจัย Age, Alcohol, Fiber, Quetelet (ตาราง 7) มีความเหมาะสมกับข้อมูลในเชิงสถิติ

ตาราง 10 ประสิทธิภาพในการพยากรณ์ของตัวแบบ

Observed	Predicted			
	Plasma Retinol Levels		Percentage	
	1	0	Correct	
Plasma Retinol Levels	1	76	59	53.1
	0	67	111	65.3
Overall Percentage			60	

ตาราง 11 การวัดความสัมพันธ์ของตัวแบบ

ตัวสถิติ	ค่าของความสัมพันธ์
Percent Concordant	66.700
Percent Discordant	32.900
c	0.669

ตาราง 12 การวัดภาวะสารูปดีของตัวแบบ

Criterion	Intercept and Covariates
AIC	413.0740
BIC	431.8050

ระดับพลาสมาเรตินอลให้ค่าร้อยละความสอดคล้อง 66.700 และค่าความสัมพันธ์แบบอื่นๆ อยู่ในระดับพอใช้ (ตาราง 11) และจากตัวแบบที่ประมาณได้ (2) สามารถพยากรณ์ภาพรวมได้ถูกต้อง 60% (ตาราง 10) โดยมีค่าน้อยกว่าของตัวแบบระดับพลาสมาเบต้าแคโรทีน อย่างไรก็ตาม เมื่อตรวจสอบจากค่าส่วนเหลือของตัวแบบพบว่า มีการกระจายตัวและมีค่าอยู่ระหว่าง [-2,2] เช่นเดียวกัน

สรุปผลการวิจัยและข้อเสนอแนะ

- ปัจจัยที่ส่งผลให้ระดับพลาสมาเบต้าแคโรทีนคือ อายุ เบต้าแคโรทีนที่บริโภคต่อวัน ดัชนีความหนาของร่างกาย เรตินอลที่บริโภคต่อวัน การสูบบุหรี่และการบริโภควิตามิน
- ปัจจัยที่ส่งผลให้ระดับพลาสมาเรตินอลคือ อายุ จำนวนครั้งของการดื่มแอลกอฮอล์ ปริมาณใยอาหารที่บริโภคต่อวันและดัชนีมวลกาย
- ตัวแบบโลจิสติก (1) ที่ประมาณได้สามารถพยากรณ์ในภาพรวมได้ถูกต้อง 81.8%
- ตัวแบบโลจิสติก (2) ที่ประมาณได้สามารถพยากรณ์ในภาพรวมได้ถูกต้อง 60.0%
- ถ้ามีการเพิ่มตัวแปรอื่น ๆ ตามที่ผู้เชี่ยวชาญได้เสนอแนะ หรือมีการควบคุมตัวแปรบางตัว อาจมีผลทำให้ตัวแบบเหมาะสมกว่าตัวแบบเดิม เช่นตัวแปรอายุ อายุที่ต่างกันถ้ากินอาหารที่มีสารเบต้าแคโรทีนเท่า ๆ กัน ระดับสารในเลือดอาจไม่แตกต่างกันก็ได้ ควรมีการศึกษาเพิ่มเติมต่อโดยควบคุมปัจจัยด้านการกินอาหารหรือวิตามินด้วย

เอกสารอ้างอิง

- สรจักร ศิริบริรักษ์. (2550). เบต้าแคโรทีน.....สารป้องกันมะเร็งมาแรง. สืบค้นเมื่อ 13 กุมภาพันธ์ 2550 จาก http://www.green-x.com/news_detail.asp?pid=307
- วีรานันท์ พงศาภักดี. (2541). การวิเคราะห์ข้อมูลเชิงกลุ่ม: ทฤษฎีและการประยุกต์ (กับ GLIM และ SPSS/FW) โรงพิมพ์มหาวิทยาลัยศิลปากร นครปฐม
- Nierenberg, D.W. and Stukel, T. (1987). Diurnal variation in plasma levels of retinol, tocopherol, and beta-carotene. *Am J Med Sci*, 294(3), 187-90
- Department of Statistics. (2001). Carnegie Mellon University, Pittsburgh, PA 15213, United States. <http://lib.stat.cmu.edu>.