

การเปรียบเทียบความแม่นยำของการศึกษาปัจจัยที่สัมพันธ์กับ
การเกิดโรคเบาหวานด้วยวิธีการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์
การถดถอยโลจิสติกพหุระดับ: ฐานข้อมูล TCNAP ของประเทศไทย ปี 2561
A Comparison of Accuracy of Factors Related to Diabetes Mellitus
by using Logistic Regression and Multilevel Logistic Regression:
TCNAP database 2018

ธนัญชัย บุญหนัก^a, พงษ์เดช สารการ^{**}

*กลุ่มวิจัยวิทยาการระบาดและป้องกันโรคมะเร็งในภูมิภาคอาเซียน

**คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

^aผู้รับผิดชอบบทความ

Tanunchai Boonnuk^a, Pongdech Sarakarn^{**}

*ASEAN Cancer Epidemiology and Prevention Research Group

**Faculty of Public Health, Khon Kaen University, Thailand

^aCorresponding author

บทคัดย่อ

ข้อจำกัดการวิเคราะห์ถดถอยโลจิสติก คือ การลดระดับชั้นเพื่อให้ตัวแปรต้นมีระดับเท่ากัน ทำให้ความละเอียดข้อมูลลดลง การวิเคราะห์ข้อมูลจึงเกิดความลำเอียง การวิจัยมีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำระหว่างการวิเคราะห์ถดถอยโลจิสติกพหุระดับและการวิเคราะห์ถดถอยโลจิสติก โดยใช้ข้อมูลสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ ปี พ.ศ. 2561 จำนวน 8,327,459 คน ผลการศึกษาพบว่า ความแตกต่างของ 95% CI ของอัตราส่วนออดส์ระหว่างทั้งสองวิธีอยู่ระหว่าง ร้อยละ 5.65 ถึง 72.14 (บางตัวแปรเกินเกณฑ์ที่กำหนดไว้คือร้อยละ 20) และความแม่นยำในภาพรวมพบว่าการวิเคราะห์การถดถอยโลจิสติกพหุระดับ (AIC=2,007,527, BIC=2,007,834, Brier score=0.0316, AUC=0.8511) ให้ผลแม่นยำกว่าการวิเคราะห์ถดถอยโลจิสติก (AIC=2,032,180, BIC=2,032,473, Brier score=0.0313, AUC=0.8505) แม้ผลเป็นเช่นนั้นแต่ค่า AUC ของทั้งสองวิธีการแตกต่างกันไม่ถึงร้อยละ 1 (น้อยกว่าเกณฑ์ที่กำหนดไว้คือร้อยละ 5) และจากการศึกษาครั้งนี้ข้อมูลไม่มีความแตกต่างระหว่างกลุ่ม (ICC=0.2656) ดังนั้นควรพิจารณาเลือกใช้การวิเคราะห์การถดถอยโลจิสติก ซึ่งข้อค้นพบที่ได้จากการศึกษาครั้งนี้นำไปสู่ข้อเสนอแนะที่ว่า หากข้อมูลมีแตกต่างระหว่างกลุ่มมาก ควรเลือกใช้สถิติการวิเคราะห์การถดถอยโลจิสติกพหุระดับ แต่หากข้อมูลไม่มีความแตกต่างระหว่างกลุ่มควรเลือกใช้สถิติการวิเคราะห์

การถดถอยโลจิสติก นอกจากนี้ควรนำข้อมูลชุดนี้ไปศึกษาเพิ่มเติมในแบบจำลองสมการโครงสร้างและการประมาณค่า

คำสำคัญ: โรคเบาหวาน ความแม่นยำ การถดถอยโลจิสติก พหุระดับ

Abstract

The limitation of traditional logistic regression analysis is that the reduction level of independent variables decrease in level of detailed data, which leads to bias in data analysis. Therefore, this study aims to evaluate and compare accuracy of traditional logistic regression and multilevel logistic regression using data from TCNAP database of the Office of Health Promotion Fund, Bureau 3, Community Health Support Office. It was observed that both technique result the different in 95% CI of odds ratio around 5.65 to 72.14% (Some variables exceeded the specified criteria, which is 20 percent.). The multilevel regression analysis (AIC=2,007,527, BIC=2,007,834, Brier score=0.0316, AUC=0.8511) gives higher accuracy compare with traditional logistic regression (AIC=2,032,180, BIC=2,032,473, Brier score=0.0313, AUC=0.8505). However, the AUC value from both techniques is less than 1% which less than standard limit 5%. Because the data used in this study have not much difference between groups (ICC=0.2656), the logistic regression can be selected. The observation from this study suggest that the multilevel regression is suitable for data that significant difference between groups, while the logistic regression is good for a small difference between groups. Nonetheless, further studies on structural equation models and model estimation are needed.

Keywords: diabetes, accuracy, logistic regression, multilevel

บทนำ

โรคไม่ติดต่อเรื้อรัง (Non-Communicable Diseases: NCDs) โดยเฉพาะโรคเบาหวาน เป็นปัญหาใหญ่ที่ทั่วโลกให้ความสำคัญ โดยโรคเบาหวานเป็นสาเหตุการเสียชีวิตอันดับ 3 ในผู้หญิง และอันดับ 8 ในผู้ชาย รวมทั้งผู้ป่วยและ

ภาครัฐต้องเสียค่าใช้จ่ายในการรักษาพยาบาล เป็นจำนวนมาก¹ นอกจากนี้สมาพันธ์เบาหวานนานาชาติ (International diabetes federation : IDF) ได้รายงานว่าเป็นปัจจุบันทั่วโลกมีผู้เสียชีวิตด้วยโรคเบาหวาน 4 ล้านคนต่อปี จากผู้เป็นเบาหวาน จำนวน 425 ล้านคน² และคาดว่าในปี

2588 จำนวนผู้ป่วยเบาหวานทั่วโลกจะเพิ่มขึ้นเป็น 629 ล้านคน³ สำหรับในประเทศไทย จากรายงานสำนักระบาดวิทยา กระทรวงสาธารณสุข ได้รายงานจำนวนผู้ป่วยโรคไม่ติดต่อเรื้อรังที่เข้ารับการรักษาในสถานพยาบาลของรัฐ ผู้ป่วยสะสมรวมจำนวน 1,219,161 คน ความชุก 1,863.39 ต่อประชากรแสนคน ผู้เสียชีวิต 12,074 ราย อัตราตาย 18.45 ต่อประชากรแสนคน โดยผู้ป่วยโรคเบาหวานที่ขึ้นทะเบียนผู้ป่วยรายใหม่ จำนวน 393,887 ราย คิดเป็นอุบัติการณ์ 602.03 ต่อประชากรแสนคน⁴

โรคเบาหวานนี้มีหลายปัจจัยเกี่ยวข้องและเชื่อมโยงกับพฤติกรรม ดังนั้นการศึกษาปัจจัยที่มีความสัมพันธ์กับการเกิดโรคจึงมีบทบาทและมีความสำคัญในการนำไปสู่การแก้ไขปัญหาและป้องกันการเกิดโรค แต่จากการศึกษาที่ผ่านมาพบว่ามีวิธีการศึกษาส่วนใหญ่ที่เกี่ยวข้องกับปัจจัยที่สัมพันธ์กับการเกิดโรคเบาหวาน เป็นการศึกษาโดยมองระดับของข้อมูลเป็นระดับเดียว เช่น งานวิจัยของ Esti Iturralde et. al.⁵ ที่ศึกษาความทุกข์ทางอารมณ์ที่เกี่ยวข้องกับโรคเบาหวาน โดยในงานวิจัยนี้ใช้ตัวแปรรายได้ครอบครัวมาวิเคราะห์ Logistic regression ร่วมกับตัวแปรระดับบุคคล และ N.C. VAZ et. al.⁶ ที่ศึกษาความชุกของโรคเบาหวานในประชากรชนบทของจังหวัดกัว ประเทศอินเดีย โดยในงานวิจัยนี้ ใช้ตัวแปรประวัติครอบครัวมาวิเคราะห์ Logistic regression ร่วมกับตัวแปรระดับบุคคล ซึ่งจากที่กล่าวมาข้างต้นจะพบว่าข้อมูลทางด้านสุขภาพโดยเฉพาะข้อมูลในระบบบริการของผู้ป่วยมีลักษณะเป็นระดับชั้น⁷ คือ ระดับครัวเรือนและระดับบุคคล โดยระดับครัวเรือน มีตัวแปรที่สัมพันธ์กับการป่วยเป็นโรคเบาหวาน ได้แก่ ลักษณะของครัวเรือน การออมเงิน และการกู้เงิน/

มีภาระหนี้สิน^{8,9} ส่วนระดับบุคคล มีตัวแปรที่สัมพันธ์กับการป่วยเป็นโรคเบาหวาน ได้แก่ เพศ อายุ ระดับการศึกษา สถานภาพสมรส อาชีพหลัก การดื่มสุรา การสูบบุหรี่ การไม่ได้พักผ่อน การไม่ได้ออกกำลังกาย การรับประทานอาหารที่ไม่เหมาะสม และการป่วยเป็นโรคความดันโลหิตสูง^{8,10,11,12,13,14,15,16,17,18,19,20} แต่วิธีการวิเคราะห์ข้อมูลที่ผ่านมาใช้วิธีแบบดั้งเดิมที่ไม่ได้พิจารณาถึงระดับของข้อมูลจนอาจทำให้เกิดความลำเอียง (bias) โดยเฉพาะการนำวิธีการวิเคราะห์ถดถอยโลจิสติกมาใช้^{5,6}

ข้อจำกัดการวิเคราะห์การถดถอยโลจิสติกคือการลดระดับชั้นเพื่อให้ตัวแปรต้นมีระดับเท่ากัน ทำให้ความละเอียดข้อมูลลดลง การวิเคราะห์ข้อมูลจึงเกิดความลำเอียง (bias)²¹ นอกจากนี้ข้อจำกัดของแบบจำลองการถดถอยระดับเดียวที่มีข้อมูลหลายระดับคือ Clustering effect ซึ่งผลของ clustering effect นี้ทำให้สูญเสียความเป็นอิสระระหว่างการวัด ดังนั้นจึงเป็นการละเมิด assumption ของการสร้างแบบจำลองการถดถอย โดยหากไม่สนใจ assumption ดังกล่าว จะทำให้การคำนวณความแปรปรวนอาจต่ำเกินไปซึ่งอาจนำไปสู่ข้อผิดพลาดชนิดที่ 1 (Type I error คือ ปฏิเสธสมมติฐาน H_0 อย่างไม่ถูกต้อง) เมื่อตรวจสอบผลการถดถอย²² ดังนั้นการนำวิธีการทางสถิติที่คำนึงถึงระดับข้อมูล เช่น การวิเคราะห์ข้อมูลระดับจึงมีความเหมาะสมในการนำมาใช้กับข้อมูลที่มีระดับแตกต่างกันสำหรับโรคไม่ติดต่อเรื้อรัง

ในการพิจารณาแบบจำลองในภาพรวมจะใช้ค่า AIC, BIC, Brier score, และ AUC โดย AIC คือเกณฑ์ที่สร้างจากการประมาณความแปรปรวนของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback-Leibler Information) ระหว่างตัวแบบจริงกับตัวแบบ

ที่เหมาะสมที่มีคุณสมบัติไม่เอนเอียง BIC คือ เกณฑ์การคัดเลือกตัวแบบที่ได้มาจากการตัดแปลงแบบเบสของเกณฑ์ AIC เรียกว่าเกณฑ์ข้อสนเทศของเบส Brier score คือ ฟังก์ชันการให้คะแนนที่เหมาะสมซึ่งวัดความถูกต้องของการคาดการณ์ความน่าจะเป็น สำหรับการทำนายแบบมิติเดียวจะเทียบเท่ากับค่าเฉลี่ยกำลังสองข้อผิดพลาดตามที่ใช้กับความน่าจะเป็นที่คาดการณ์ไว้ และ AUC คือพื้นที่ใต้โค้งเส้นกราฟ ROC

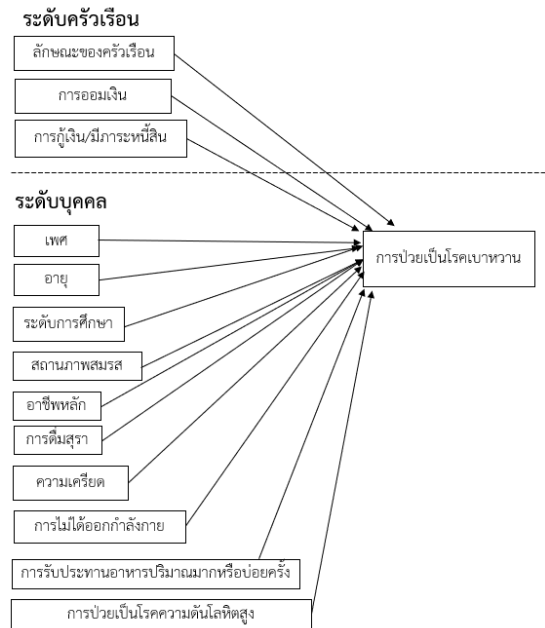
การศึกษาครั้งนี้จึงต้องการเปรียบเทียบให้เห็นความแตกต่างที่เกิดขึ้นโดยการวิเคราะห์หาปัจจัยที่มีสัมพันธ์กับการเกิดโรคเบาหวานด้วยวิธีการ Logistic regression และ Multilevel logistic regression โดยการวิเคราะห์ในรายตัวแปรด้วย 95% CI of Odds ratio วิเคราะห์ภาพรวมโดยการหาค่า AIC, BIC, Brier score, และ AUC และเปรียบเทียบระหว่าง 2 วิธีการ โดยใช้ข้อมูลจากฐานข้อมูล TCNAP ของสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.) สำนัก 3 สำนักสนับสนุนสุขภาวะชุมชน ในปี พ.ศ. 2561 เพื่อประเมินความเหมาะสมกับลักษณะข้อมูลดังกล่าวข้างต้น โดยหากพบว่ามี ความแตกต่างระหว่างกลุ่ม (พิจารณาจากค่า ICC) ควรหลีกเลี่ยงไปใช้วิธีการ Multilevel logistic regression แต่หากพบที่ไม่มี ความแตกต่างระหว่างกลุ่ม ก็ยังคงสามารถใช้ Logistic regression ได้

วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบความแม่นยำของผลลัพธ์ที่ได้ระหว่าง วิธีการ Logistic regression และวิธีการ Multilevel logistic regression

กรอบแนวความคิดของการวิจัย

จากการทบทวนวรรณกรรมและรายงานการวิจัยที่เกี่ยวข้อง การศึกษาครั้งนี้มีกรอบแนวความคิดของการวิจัยดังรูปที่ 1



รูปที่ 1 กรอบแนวความคิดของการวิจัย

วิธีการศึกษา

รูปแบบการวิจัย การศึกษานี้เป็นการใช้ข้อมูลทุติยภูมิจากฐานข้อมูล TCNAP ของ สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.) สำนัก 3 สำนักสนับสนุนสุขภาวะชุมชน ภายใต้แนวคิดของหลักการ Data mining มาใช้ในการวิเคราะห์ Mathematical model

ประชากรในการศึกษาครั้งนี้ได้แก่ ประชาชนที่ถูกเก็บรวบรวมข้อมูลอยู่ในฐานข้อมูล TCNAP ของสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.) สำนัก 3 สำนักสนับสนุนสุขภาวะชุมชน ในปี พ.ศ. 2561 โดยเริ่มต้นจากข้อมูลจากฐานข้อมูลทั้งหมด จำนวน 8,485,838 คน (2,393,032

ครัวเรือน) ถูกคัดออกเนื่องจาก อายุ น้อยกว่า 0 หรือมากกว่า 120 ปี จำนวน 12,016 คน (10,643 ครัวเรือน) และถูกคัดออกเนื่องจากข้อมูลไม่สมบูรณ์ครบทุกตัวแปร จำนวน 146,363 คน (19,740 ครัวเรือน) ทำให้เหลือข้อมูลที่ถูกนำไปใช้ในการวิจัยครั้งนี้จำนวน 8,327,459 คน (2,362,649 ครัวเรือน)

กลุ่มตัวอย่าง ที่นำมาพิจารณาในการประเมินตัวแบบครั้งนี้ จำแนกออกเป็น 2 กลุ่ม คือ 1) ตัวอย่างที่นำมาพิจารณาเพื่อพัฒนาตัวแบบ ได้แก่ ตัวอย่างที่ถูกสุ่มมาแบบง่ายโดยใช้โปรแกรม STATA ร้อยละ 80 จากตัวอย่างข้างต้น 2) ตัวอย่างที่นำมาพิจารณาเพื่อประเมินตัวแบบ ได้แก่ ตัวอย่างที่ถูกสุ่มมาแบบง่ายโดยใช้โปรแกรม STATA ร้อยละ 20 จากตัวอย่างข้างต้น (เกณฑ์ 80:20 อ้างอิงจาก John Neumann23)

แหล่งข้อมูลที่ใช้ในการวิจัย

ใช้ข้อมูลจากฐานข้อมูล TCNAP ของสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.) สำนัก 3 สำนักสนับสนุนสุขภาวะชุมชน โดยเป็นข้อมูลในปี พ.ศ. 2561 โดยข้อมูลถูกเก็บโดยองค์กรปกครองส่วนท้องถิ่นที่มีความร่วมมือ โดยข้อมูลที่เก็บเป็นข้อมูล ณ ปี 2561

การเก็บรวบรวมข้อมูล

ในการศึกษาครั้งนี้ ได้จำแนกวิธีการเก็บรวบรวมข้อมูลออกเป็นขั้นตอน ดังต่อไปนี้

ขั้นตอนการเตรียมการ ศึกษาทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับ multilevel logistic regression

ขั้นตอนดำเนินงาน ติดต่อบริษัทงานเพื่อขอข้อมูลจากฐานข้อมูล TCNAP ในปี พ.ศ. 2561

ของสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.) สำนัก 3 สำนักสนับสนุนสุขภาวะชุมชน

ขั้นตอนตรวจสอบความถูกต้องของข้อมูล

1) ทำการแจกแจงความถี่ในตัวแปรประเภทแจกแจงว่ามีค่าที่ผิดปกติหรือไม่ 2) พิจารณาค่าต่ำสุด สูงสุด ในตัวแปรต่อเนื่องว่ามีค่าที่ผิดปกติหรือไม่ เช่น อายุ ต้องไม่ติดลบและไม่ควรเกิน 120 โดยหากมีค่าที่ผิดปกติ ให้ทำการแปลงค่าดังกล่าวให้เป็น missing value 3) พิจารณาความสอดคล้องของข้อมูลโดยใช้ผู้ตรวจสอบความถูกต้องจำนวน 2 คน และแยกกันปฏิบัติข้อ 1) และ 2) อย่างเป็นอิสระต่อกัน แล้วจึงนำข้อมูลของทั้ง 2 คน มาตรวจสอบว่าตรงกันหรือไม่

การพิทักษ์สิทธิกลุ่มตัวอย่าง

การวิจัยครั้งนี้ได้ผ่านการพิจารณาจริยธรรมการวิจัยจากศูนย์จริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยขอนแก่น รหัสเลขที่ HE632294 ลงวันที่ 16 ธันวาคม 2563 โดยข้อมูลทุกอย่างที่ได้จากการศึกษาถูกเก็บเป็นความลับ มีเพียงผู้วิจัยเท่านั้นที่ได้อ่านและรับทราบข้อมูล และนำเสนอข้อมูลในลักษณะภาพรวม

การวิเคราะห์ข้อมูล

1) การวิเคราะห์ข้อมูลบุคคล และข้อมูลครัวเรือน โดยใช้สถิติเชิงพรรณนา กรณีสอดคล้องกับจำนวนและร้อยละ กรณีสอดคล้องกับค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่ามัธยฐาน ค่าต่ำสุด และค่าสูงสุด และทำการจัดกลุ่มเพื่อแสดงจำนวนและร้อยละ

2) การวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างปัจจัยทั้งในระดับบุคคลและครัวเรือนกับ

การป่วยเป็นโรคเบาหวาน ใช้สถิติ logistic regression และ multilevel logistic regression (mixed effect logistic regression) โดยหาค่าความเชื่อมั่นที่ระดับ 95% ของค่าสัมประสิทธิ์ถดถอยรายปัจจัย และประเมินภาพรวมของตัวแบบด้วย AIC, BIC, Brier score และ AUC (Area under ROC curve) และเพื่อทำการเปรียบเทียบความแม่นยำของทั้ง 2 วิธี โดยในการวิเคราะห์ข้อมูลมีการคำนึงถึง

over fit model, under fit model, และ misspecified model โดยพิจารณาการนำเข้าข้อมูลรายตัวแปรจากสถิติทดสอบ LR test นอกจากนี้ในการทดสอบเรื่องของ Multicollinearity โดยใช้ค่า VIF พบว่าค่า VIF อยู่ระหว่าง 1.03 - 2.71 ซึ่งน้อยกว่า 10 จึงสรุปได้ว่าไม่มีปัญหาในเรื่อง Multicollinearity

3) สำหรับร้อยละความต่างคำนวณดังนี้

$$\Delta\%adjOR = 100x \frac{adjOR_{MLR} - adjOR_{TLR}}{adjOR_{TLR}}$$

$$\Delta\%range = 100x \frac{(\text{range of 95\%CI of OR})_{MLR} - (\text{range of 95\%CI of OR})_{TLR}}{(\text{range of 95\%CI of OR})_{TLR}}$$

$$\Delta\%AUC = 100x \frac{AUC_{MLR} - AUC_{TLR}}{AUC_{TLR}}$$

เมื่อ $\Delta\%$ คือ ร้อยละผลต่าง, range คือ พิสัย, OR คือ Odds ratio, adjOR คือ Adjusted Odds ratio, AUC คือ Area under ROC curve, MLR คือ การวิเคราะห์การถดถอยโลจิสติกพหุระดับ (Multilevel logistic regression) และ “TLR” คือ การวิเคราะห์การถดถอยโลจิสติก (Traditional logistic regression)

การเปรียบเทียบความแตกต่างรายตัวแปรระหว่างทั้งสองวิธี พิจารณาจาก $\Delta\%range$ หากเกินร้อยละ 20 แสดงว่าทั้ง 2 วิธีการให้ผลที่แตกต่างกัน (พัฒนาจากงานวิจัยของ Brayan Alexander Fonseca Martinez and Vanessa Bielefeldt Leotti โดยเปลี่ยนจาก Point estimate เป็นช่วงความเชื่อมั่น)

การเปรียบเทียบความแตกต่างภาพรวมระหว่างทั้งสองวิธี พิจารณาจาก $\Delta\%AUC$ หากเกินร้อยละ 5 แสดงว่าทั้ง 2 วิธีการให้ผลที่แตกต่างกัน

ผลการศึกษา

ข้อมูลเชิงพรรณนา

1) กลุ่มตัวอย่างส่วนใหญ่เป็นเพศหญิง ร้อยละ 50.77 มีอายุ 35 ถึง 59 ปี ร้อยละ 39.69 (อายุเฉลี่ย 41.62 ปี) จบการศึกษาระดับประถมศึกษา ร้อยละ 46.48 สมรส/อยู่ด้วยกันโดยไม่จดทะเบียน ร้อยละ 52.53 และไม่ได้ประกอบอาชีพ ร้อยละ 33.26

2) กลุ่มตัวอย่างส่วนใหญ่เป็นไม่ดื่มสุรา ร้อยละ 92.08 ไม่เครียด ร้อยละ 98.42 ออกกำลังกาย ร้อยละ 94.85 ไม่ได้รับประทานอาหารปริมาณมากหรือบ่อยครั้ง ร้อยละ 98.47 และไม่ได้ป่วยเป็นโรคความดันโลหิตสูง ร้อยละ 93.42

3) กลุ่มตัวอย่างส่วนใหญ่มีจำนวนสมาชิกในครัวเรือนไม่เกิน 5 คน ร้อยละ 85.42 (จำนวนสมาชิกในครัวเรือนเฉลี่ย 3.52 คน/ครัวเรือน) เป็นครอบครัวขยาย ร้อยละ 78.32 ไม่มีการออมเงิน

ร้อยละ 52.08 และกู้เงินหรือมีภาระหนี้สิน ร้อยละ 62.23

4) กลุ่มตัวอย่างส่วนใหญ่ไม่ได้ป่วยเป็นโรคเบาหวานร้อยละ 96.45

ค่า 95% CI ของ Odds ratio

ในการเปรียบเทียบ 95% CI of Odds ratio ระหว่างการวิเคราะห์ถดถอยโลจิสติกพหุระดับ และการวิเคราะห์ถดถอยโลจิสติก โดยพิจารณาจากค่า $\Delta\%range$ พบว่ามีค่าระหว่างร้อยละ 5.65 ถึง 72.14 โดยตัวแปรที่มีความแตกต่างเกินเกณฑ์ที่กำหนด ($\Delta\%range > 20$) ได้แก่ Intercept

อายุตั้งแต่ 60 ปีขึ้นไป อาชีพเกษตรกร ความเครียด การไม่ได้ออกกำลังกาย การรับประทานอาหารปริมาณมากหรือบ่อยครั้ง การป่วยเป็นโรคความดันโลหิตสูง ประเภทครัวเรือน การออมเงิน และการกู้เงิน/มีภาระหนี้สิน และพบตัวแปรที่มีความแตกต่างกัน แต่ยังไม่ถึงเกณฑ์ที่ยอมรับ ($\Delta\%range \leq 20$) ได้แก่ เพศ อายุ 35-59 ปี ระดับการศึกษา สถานภาพสมรส อาชีพรับจ้าง/ลูกจ้าง อาชีพเจ้าของกิจการ/กิจการของครอบครัว อาชีพรับราชการ/รัฐวิสาหกิจ และการดื่มสุรา รายละเอียดดังตารางที่ 1

ตารางที่ 1 การเปรียบเทียบ 95% CI ของ Odds ratio ระหว่างการถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับ

ตัวแปร		จำนวน	Multilevel logistic regression			Logistic regression			$\Delta\%$ adjOR	$\Delta\%$ range		
			Adj. OR	95% CI		SE	Adj. OR	95% CI			SE	
Intercept		8,327,459	0.0017**	0.0017	0.0018	<0.0001	0.0041**	0.0040	0.0042	<0.0001	-57.28	-45.05
เพศ	ชาย	4,099,848	1.0000				1.0000					
	หญิง	4,227,611	1.5491**	1.5349	1.5635	0.0073	1.4669**	1.4546	1.4792	0.0063	5.61	16.06
อายุ	ต่ำกว่า 35 ปี	3,224,519	1.0000				1.0000					
	35-59 ปี	3,305,350	3.5450**	3.4844	3.6066	0.0311	3.3836**	3.3289	3.4393	0.0282	4.77	10.57
	ตั้งแต่ 60 ปี ขึ้นไป	1,797,590	8.9910**	8.8378	9.1469	0.0789	7.4751**	7.3567	7.5953	0.0609	20.28	29.53
ระดับการศึกษา	ไม่ได้เรียน/ยังไม่สำเร็จการศึกษา	1,029,151	1.0000				1.0000					
	ประถมศึกษา	3,870,892	1.3763**	1.3558	1.3971	0.0105	1.3256**	1.3085	1.3430	0.0088	3.82	19.88
	มัธยมศึกษา/อนุปริญญา	2,771,532	0.8980**	0.8817	0.9145	0.0084	0.9176**	0.9029	0.9325	0.0076	-2.14	10.48
	ตั้งแต่ปริญญาตรีขึ้นไป	655,884	0.6988**	0.6786	0.7197	0.0105	0.7315**	0.7123	0.7512	0.0099	-4.46	5.65
สถานภาพสมรส	หม้าย/หย่า/แยกกันอยู่	1,358,607	1.0000				1.0000					
	โสด	2,594,777	1.0482**	1.0355	1.0611	0.0065	1.0602**	1.0490	1.0715	0.0057	-1.13	13.92
	สมรส/อยู่ด้วยกันโดยไม่จดทะเบียน	4,374,075	0.8160**	0.8025	0.8298	0.0069	0.8499**	0.8375	0.8624	0.0064	-3.98	9.22
อาชีพหลัก	ไม่ได้ประกอบอาชีพ	2,769,921	1.0000				1.0000					
	เกษตรกร	2,414,418	1.1160**	1.1027	1.1295	0.0069	1.0762**	1.0651	1.0874	0.0057	3.70	20.51
	รับจ้าง/ลูกจ้าง	2,359,671	0.6974**	0.6876	0.7073	0.0050	0.7229**	0.7140	0.7319	0.0046	-3.53	9.55
	เจ้าของกิจการ/กิจการของครอบครัว	521,662	1.0882**	1.0682	1.1085	0.0103	1.0727**	1.0557	1.0900	0.0087	1.44	17.60
	รับราชการ/รัฐวิสาหกิจ	261,787	1.0707**	1.0366	1.1059	0.0177	1.0683**	1.0376	1.0999	0.0159	0.22	11.14
การดื่มสุรา	ไม่ดื่มสุรา	7,667,964	1.0000				1.0000					
	ดื่มสุรา	659,495	1.0140	0.9972	1.0311	0.0086	1.0236**	1.0085	1.0388	0.0077	-0.94	11.93

ตัวแปร	จำนวน	Multilevel logistic regression				Logistic regression			Δ% adjOR	Δ% range	
		Adj. OR	95% CI		SE	Adj. OR	95% CI				SE
ความเครียด	ไม่เครียด	8,195,740	1.0000				1.0000				
	เครียด	131,719	1.5755**	1.5392	1.6127	0.0187	1.4341**	1.4064	1.4623	0.0143	9.86
การไม่ได้ ออกกำลังกาย	ออกกำลังกาย	7,898,258	1.0000				1.0000				
	ไม่ได้ออกกำลังกาย	429,201	1.5663**	1.5419	1.5911	0.0125	1.4331**	1.4145	1.4519	0.0095	9.30
การรับประทานอาหาร ปริมาณ มาก	ไม่ใช้	8,200,263	1.0000				1.0000				
	ใช้	127,196	1.5565**	1.5096	1.6049	0.0243	1.4868**	1.4487	1.5260	0.0197	4.69
การป่วยเป็นโรค ความดันโลหิตสูง	ไม่ป่วย	7,779,195	1.0000				1.0000				
	ป่วย	548,264	9.4250**	9.3212	9.5300	0.0533	7.0425**	6.9821	7.1034	0.0310	33.83
ประเภท ครัวเรือน	ครอบครัวขยาย	7,643,795	1.0000				1.0000				
	ครอบครัวเดี่ยว	683,664	1.1016**	1.0864	1.1170	0.0078	1.0855**	1.0729	1.0983	0.0065	1.48
การออมเงิน	ไม่มีการออมเงิน	4,158,296	1.0000				1.0000				
	มีการออมเงิน	4,169,163	1.0675**	1.0570	1.0781	0.0054	1.0489**	1.0404	1.0575	0.0044	1.77
การกู้เงิน/ มีภาระหนี้สิน	ไม่ได้กู้/ไม่มีภาระหนี้สิน	2,647,079	1.0000				1.0000				
	กู้เงิน/มีภาระหนี้สิน	5,680,380	1.2344**	1.2210	1.2480	0.0069	1.1942**	1.1834	1.2050	0.0055	3.37

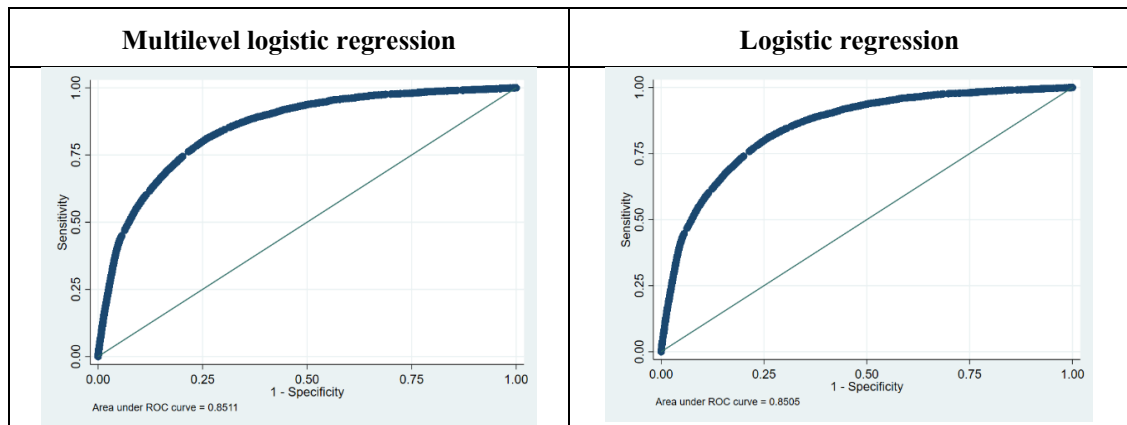
หมายเหตุ ** หมายถึง $p < .01$, * หมายถึง $p < .05$

ค่าความแม่นยำในภาพรวมของแบบจำลอง

จากการศึกษา พบว่า การวิเคราะห์การถดถอยโลจิสติกพหุระดับ จะให้ค่า AIC และ BIC ที่น้อยกว่าการวิเคราะห์การถดถอยโลจิสติก และเมื่อพิจารณาค่า Brier score และ AUC (Area under ROC curve) แล้วพบว่า การวิเคราะห์การถดถอยโลจิสติกพหุระดับ จะให้ค่าดังกล่าวที่สูงกว่าการวิเคราะห์การถดถอยโลจิสติก ดังนั้นทุกค่าจึงบ่งชี้ไปในทิศทางเดียวกันว่า ในภาพรวมของแบบจำลอง การวิเคราะห์การถดถอยโลจิสติกพหุระดับจะมีความแม่นยำสูงกว่าการวิเคราะห์การถดถอยโลจิสติก ซึ่งแม้ผลเป็นเช่นนั้นแต่ค่า AUC ของทั้งสองวิธีการแตกต่างกันไม่ถึงร้อยละ 1 (น้อยกว่าเกณฑ์ที่กำหนดไว้คือร้อยละ 5)

นอกจากนี้จากการศึกษาข้อมูลที่มีขนาดตัวอย่างกันคือ 1,000 5,000 10,000 50,000 100,000 500,000 1,000,000 5,000,000 และ 8,327,459

พบว่า เมื่อข้อมูลตั้งแต่ 5,000 คน ขึ้นไป จะให้ค่า AUC ใกล้เคียงกันและมีค่ามากกว่า 0.80 และเมื่อพิจารณาค่า ICC ในแต่ละขนาดตัวอย่าง พบว่าค่า ICC มีค่าน้อยกว่าเกณฑ์ที่กำหนด แสดงให้เห็นว่าข้อมูลไม่มีความแตกต่างระหว่างกลุ่ม ทำให้ผลการวิเคราะห์ของทั้ง 2 วิธีการใกล้เคียงกันมาก จึงควรเลือกใช้สถิติการวิเคราะห์การถดถอยโลจิสติก รายละเอียดดังรูปที่ 2 และตารางที่ 2



รูปที่ 2 ROC และ AUC (Area under ROC curve) ระหว่างการถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับ

ตารางที่ 2 การเปรียบเทียบภาพรวมของแบบจำลองระหว่างการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับ ในแต่ละขนาดตัวอย่าง

ขนาดตัวอย่าง	Multilevel logistic regression					Logistic regression				Δ% AUC
	ICC	AIC	BIC	Brier Score	AUC	AIC	BIC	Brier Score	AUC	
1,000	0.4580	218.89	314.53	0.0327	0.5503	218.05	308.91	0.0327	0.5503	0.00
5,000	0.3133	1,116.54	1,259.91	0.0340	0.8501	1,131.09	1,267.95	0.0329	0.8491	0.12
10,000	0.2772	2,287.37	2,446.00	0.0352	0.8624	2,304.66	2,456.08	0.0348	0.8603	0.24
50,000	0.2619	12,034.66	12,228.69	0.0318	0.8537	12,159.31	12,344.53	0.0315	0.8525	0.14
100,000	0.2454	24,248.40	24,457.68	0.0316	0.8499	24,453.54	24,653.31	0.0313	0.8494	0.06
500,000	0.2637	120,441.00	120,685.70	0.0316	0.8507	121,745.60	121,979.10	0.0312	0.8502	0.06
1,000,000	0.2616	241,272.10	241,532.00	0.0315	0.8536	243,940.70	244,188.80	0.0312	0.8530	0.07
5,000,000	0.2673	1,203,483	1,203,778	0.0313	0.8509	1,218,707	1,218,989	0.0309	0.8503	0.07
8,327,459	0.2656	2,007,527.00	2,007,834.00	0.0316	0.8511	2,032,180.00	2,032,473.00	0.0313	0.8505	0.07

หมายเหตุ 1) ICC<0.5 (Terry K. Koo, and Mae Y. Li, 2016)²⁵

2) ขนาดตัวอย่าง 500 ไม่สามารถทำได้เนื่องจากข้อมูลไม่สมบูรณ์ครบทุกตัวแปร

อภิปรายผลการวิจัย

จากผลการวิจัย ผู้วิจัยอภิปรายผลในแต่ละประเด็น ดังนี้

การอภิปรายผลปัจจัยที่มีสัมพันธ์กับการเกิดโรคเบาหวาน

1) เพศ มีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่าเพศหญิงมีความเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าเพศชาย ($p < .05$) ซึ่งสอดคล้องกับผลการวิจัยของ Ashis Talukder and Zobayer Hossain⁸ ทั้งนี้อาจเป็นเพราะเพศหญิงโดยส่วนใหญ่มีพฤติกรรมการรับประทานอาหารในปริมาณมาก และบ่อยครั้งและประเภทอาหารที่มีปริมาณน้ำตาลสูง จึงมีความเสี่ยงต่อการป่วยเป็นเบาหวานมากกว่าเพศชาย

2) อายุมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่าเมื่ออายุมากขึ้น ความเสี่ยงต่อการเกิดโรคเบาหวานก็จะสูงขึ้นเช่นกัน ($p < .05$) ซึ่งสอดคล้องกับผลการวิจัยของสุดสงวน ชินโน¹⁷ ทั้งนี้อาจเป็นเพราะเมื่ออายุมากขึ้น ระบบการเผาผลาญพลังงานของร่างกายก็จะลดลง ทำให้โอกาสเกิดโรคเบาหวานก็จะมากขึ้นเช่นกัน

3) ระดับการศึกษามีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ระดับการศึกษาประถมศึกษา มีความเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าไม่ได้เรียน/ยังไม่สำเร็จการศึกษา ($p < .05$) ในขณะที่ระดับการศึกษามัธยมศึกษา/อนุปริญญา และตั้งแต่ปริญญาตรีขึ้นไปมีความเสี่ยงต่อการเกิดโรคเบาหวานน้อยกว่าไม่ได้เรียน/ยังไม่สำเร็จการศึกษา ($p < .05$) ซึ่งขัดแย้งกับผลการวิจัยของ Ashis Talukder and Zobayer Hossain⁸ ทั้งนี้ อาจเป็นเพราะเมื่อระดับการศึกษาเพิ่มสูงขึ้น ความรู้เกี่ยวกับโรคเบาหวานก็เพิ่มมากขึ้น ทำให้

มีพฤติกรรมการป้องกันโรคเบาหวานปรับเปลี่ยน จึงทำให้การป่วยเป็นโรคเบาหวานลดลง

4) สถานภาพสมรสมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ผู้ที่มีสถานภาพโสด มีความเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าหม้าย/หย่า/แยกกันอยู่ ($p < .05$) ในขณะที่ผู้ที่มีสถานภาพสมรส/อยู่ด้วยกันโดยไม่จดทะเบียนมีความเสี่ยงต่อการเกิดโรคเบาหวานน้อยกว่าหม้าย/หย่า/แยกกันอยู่ ($p < .05$) ซึ่งสอดคล้องกับผลการวิจัยของ จารุพร สฤกษ์สุข¹⁵ ทั้งนี้อาจเป็นเพราะผู้ที่โสด อาจไม่ได้มีผู้ที่คอยควบคุมหรือเตือนในเรื่องพฤติกรรมการรับประทานอาหารและพฤติกรรมสุขภาพ ต่อการป้องกันโรคเบาหวาน (เช่น การออกกำลังกาย) ในขณะที่ผู้ที่สมรสแล้วอาจมีผู้คอยตักเตือนหรือควบคุมพฤติกรรมการควบคุมป้องกันโรคดังกล่าว

5) อาชีพหลักมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ผู้ที่ประกอบอาชีพเกษตรกร เจ้าของกิจการ/กิจการของครอบครัว รับราชการ/รัฐวิสาหกิจมีความเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าผู้ที่ไม่ได้ประกอบอาชีพ ($p < .05$) ในขณะที่ผู้ประกอบอาชีพรับจ้าง/ลูกจ้าง มีความเสี่ยงต่อการเกิดโรคเบาหวานน้อยกว่าผู้ที่ไม่ได้ประกอบอาชีพ ($p < .05$) ซึ่งสอดคล้องกับงานวิจัยของชานาญ สูญสันภัย¹⁶ ทั้งนี้อาจเป็นเพราะที่ประกอบอาชีพเกษตรกร เจ้าของกิจการ/กิจการของครอบครัว รับราชการ/รัฐวิสาหกิจ ส่วนใหญ่มีรายได้ที่เพียงพอที่จะจับจ่ายซื้ออาหารมารับประทาน โดยเฉพาะอาหารที่มีปริมาณน้ำตาลสูง ในขณะที่ผู้ประกอบอาชีพรับจ้าง/ลูกจ้างส่วนใหญ่ไม่ได้มีรายได้ที่เพียงพอที่จะจับจ่ายซื้ออาหารมารับประทาน จึงรับประทานอาหารที่มีปริมาณน้ำตาลต่ำกว่า

6) ความเครียดมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ผู้ที่มีความเครียดจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าผู้ที่ไม่มีความเครียด ($p < .05$) ซึ่งสอดคล้องกับงานวิจัยของกัณทิมาอยู่รวม¹² ทั้งนี้อาจเป็นเพราะผู้ที่มีความเครียดจะไม่ได้มีเวลาดูแลพฤติกรรมสุขภาพของตน อีกทั้งยังมีการรับประทานในปริมาณมากและบ่อยครั้งเพื่อบรรเทาความเครียด จึงทำให้มีความเสี่ยงต่อการเกิดโรคเบาหวานที่สูงกว่า

7) การไม่ได้ออกกำลังกายมีความสัมพันธ์กับการเกิดโรคเบาหวาน ($p < .05$) โดยพบว่า ผู้ที่ไม่ได้ออกกำลังกายจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าผู้ที่ออกกำลังกาย ซึ่งสอดคล้องกับงานวิจัยวิจิต อินทร์ลำพันธ์¹⁸ ทั้งนี้อาจเป็นเพราะผู้ที่ไม่ออกกำลังกายเป็นประจำ จะทำให้เกิดการเผาผลาญพลังงานของร่างกาย ส่งผลให้ระดับน้ำตาลในเลือดลดลง

8) การรับประทานอาหารปริมาณมากหรือบ่อยครั้งมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ผู้ที่การรับประทานอาหารปริมาณมากหรือบ่อยครั้งจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าผู้ที่ไม่ได้รับประทานอาหารปริมาณมากหรือบ่อยครั้ง ($p < .05$) ซึ่งสอดคล้องกับงานวิจัยของอนัญญา ประดิษฐ์ปรีชา¹⁴ ทั้งนี้อาจเป็นเพราะการรับประทานอาหารปริมาณมากหรือบ่อยครั้งจะทำให้ระดับน้ำตาลในเลือดเพิ่มสูงขึ้น ส่งผลให้ป่วยเป็นโรคเบาหวาน

9) ป่วยเป็นโรคความดันโลหิตสูงมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ผู้ที่ป่วยเป็นโรคความดันโลหิตสูงจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าผู้ที่ไม่ได้ป่วยเป็นโรคความดันโลหิตสูง ($p < .05$) ซึ่งสอดคล้องกับงานวิจัยของวิจิต

อินทร์ลำพันธ์¹⁸ ทั้งนี้อาจเป็นเพราะโรคเบาหวานและโรคความดันโลหิตสูงมักเป็นโรคที่มาพร้อมกันที่โดยส่วนใหญ่จะเกิดกับผู้สูงอายุที่ระบบการเผาผลาญร่างกายน้อยกว่ากลุ่มวัยรุ่นและวัยแรงงาน

10) ประเภทครัวเรือนมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ครอบครัวเดี่ยวจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าครอบครัวขยาย ($p < .05$) ซึ่งสอดคล้องกับงานวิจัยของ Randa M. Kutob et. al.⁹ ทั้งนี้อาจเป็นเพราะครอบครัวขยายจะมีผู้ที่คอยช่วยแทรกแซงพฤติกรรมสุขภาพที่ไม่เหมาะสมที่เสี่ยงต่อการป่วยเป็นโรคเบาหวาน ในขณะที่ครอบครัวเดี่ยวอาจไม่มีสมาชิกอื่นที่มาช่วยแทรกแซงพฤติกรรมสุขภาพที่ไม่เหมาะสม ทำให้มีพฤติกรรมที่นำไปสู่การป่วยเป็นโรคเบาหวาน

11) การออมเงินมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ครอบครัวที่มีการออมเงินจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าครอบครัวที่ไม่ได้มีการออมเงิน ($p < .05$) ซึ่งสอดคล้องกับงานวิจัยของ Ashis Talukder and Zobayer Hossain⁸ ทั้งนี้อาจเป็นเพราะการออมเงินบ่งบอกถึงสถานะทางเศรษฐกิจของครอบครัว ว่ามีเงินเพียงพอต่อการใช้จ่ายใช้สอย ทำให้ส่งผลต่อการเอื้อต่อการใช้จ่ายเพื่อการรับประทานอาหารที่ไม่เหมาะสมและนำไปสู่ความเสี่ยงต่อการป่วยเป็นโรคเบาหวาน

12) การกู้เงิน/มีภาระหนี้สินมีความสัมพันธ์กับการเกิดโรคเบาหวาน โดยพบว่า ครอบครัวที่มีการกู้เงิน/มีภาระหนี้สินจะเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าครอบครัวที่ไม่ได้มีการกู้เงิน/มีภาระหนี้สิน ($p < .05$) ซึ่งขัดแย้งกับงานวิจัยของ Ashis Talukder and Zobayer Hossain⁸ ทั้งนี้อาจเป็นเพราะอาจ ในครอบครัวมีการกู้เงิน/มีภาระ

หนังสือเพื่อมาใช้ในการจับจ่ายเพื่อส่งเสริมพฤติกรรมทางสุขภาพที่ไม่เหมาะสม อีกทั้งการเป็นหนี้ก็ส่งผลให้เกิดภาวะเครียดและนำไปสู่พฤติกรรมมารับประทานอาหารปริมาณมากและเสี่ยงต่อการเกิดโรคเบาหวานได้เช่นกัน

13) การดื่มสุรา โดยจากการศึกษาพบข้อขัดแย้งในเรื่องความสัมพันธ์ระหว่างการดื่มสุราและการเกิดโรคเบาหวานระหว่างการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับ โดยการวิเคราะห์การถดถอยโลจิสติกพบว่า การดื่มสุรามีความสัมพันธ์กับการเกิดโรคเบาหวานโดยผู้ที่ดื่มสุราจะมีความเสี่ยงต่อการเกิดโรคเบาหวานสูงกว่าผู้ที่ไม่ได้ดื่มสุรา ($p < .05$) ในขณะที่การวิเคราะห์การถดถอยโลจิสติกพหุระดับกลับไม่พบความสัมพันธ์ระหว่างการดื่มสุราและการเกิดโรคเบาหวานดังกล่าว ($p > .05$) ซึ่งทั้งสองคล่องและขัดแย้งกับงานวิจัยของ S. Carlsson et. al. (2005)²⁰ ทั้งนี้อาจเป็นเพราะการวิเคราะห์การถดถอยโลจิสติกพหุระดับจะพิจารณาถึงลำดับชั้นในขณะที่การวิเคราะห์การถดถอยโลจิสติก จะพิจารณาเพียงตัวบุคคลเท่านั้น ทำให้ผลที่ออกมาในบางตัวแปรมีความแตกต่างกัน

การอภิปรายผลการเปรียบเทียบค่า 95% CI of Odds ratio ระหว่างการถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับ

จากผลการวิจัยที่พบว่าค่า $\Delta\%$ range อยู่ระหว่างร้อยละ 5.65 ถึง 72.14 ซึ่งมีบางตัวแปรที่มีค่า $\Delta\%$ range เกินเกณฑ์ที่กำหนด คือร้อยละ 20 ทั้งนี้อาจเนื่องมาจากขนาดตัวอย่างมีขนาดใหญ่ ทำให้ช่วง 95% CI แคบ และถึงแม้จะมี Standard error (SE) แต่ค่าที่มีค่อนข้างน้อย ทั้งนี้เนื่องจาก Standard error ถูกพิจารณาจากการหารด้วย

ขนาดตัวอย่าง เพราะฉะนั้น เมื่อขนาดตัวอย่างโต ค่า Standard error จึงน้อย พอค่า Standard error น้อย จึงทำให้ค่า 95% CI แคบเข้า ดังนั้นในการศึกษาครั้งนี้เมื่อขนาดตัวอย่างมีขนาดใหญ่ จึงทำให้มองเห็นภาพความแตกต่างไม่ชัดเจน แต่ในทางปฏิบัติการวิจัยโดยทั่ว ๆ ไป มักใช้ขนาดตัวอย่างที่น้อย จึงทำให้มองเห็นความแตกต่างระหว่างการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับได้ชัดเจน ดังนั้นจึงพอสรุปได้ว่าเมื่อขนาดตัวอย่างมีขนาดใหญ่ การวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับจะไม่มี ความแตกต่างกันหรือแตกต่างกันอย่างไม่ชัดเจน

การอภิปรายผลการเปรียบเทียบค่า AIC BIC Brier score และ AUC (Area under ROC curve) ระหว่างการถดถอยโลจิสติกและการวิเคราะห์การถดถอยโลจิสติกพหุระดับ

ทุกค่าจึงบ่งชี้ไปในทิศทางเดียวกันว่า ในภาพรวมของแบบจำลอง การวิเคราะห์การถดถอยโลจิสติกพหุระดับจะมีความแม่นยำสูงกว่าการวิเคราะห์การถดถอยโลจิสติก ซึ่งสอดคล้องกับงานวิจัยของ Jindra Reissigová et. al.²⁴ ทั้งนี้ อาจเป็นเพราะค่า 95% CI ของ Odds ratio ของการวิเคราะห์การวิเคราะห์การถดถอยโลจิสติกพหุระดับจะมีมิติในการคำนวณที่สูงกว่า คือพิจารณาลำดับชั้นด้วย ทำให้ค่า 95% CI กว้างกว่า ในขณะที่ในเรื่องความแม่นยำของทั้งแบบจำลอง เมื่อนำลำดับชั้นมาพิจารณาด้วย ค่าความแม่นยำของวิธีการวิเคราะห์การถดถอยโลจิสติกพหุระดับจึงมีค่ามากกว่าวิธีการวิเคราะห์การถดถอยโลจิสติก ซึ่งแม้ผลเป็นเช่นนั้นแต่ค่า AUC ของทั้งสองวิธีการแตกต่างกันไม่ถึงร้อยละ 1 (น้อยกว่าเกณฑ์ที่กำหนดไว้คือร้อยละ 5) โดย

ความไม่แตกต่างกันของทั้ง 2 วิธีการนี้ สอดคล้องกับงานวิจัยของ Roxana Alexandrescu et. al.²⁶ นอกจากนี้เนื่องจากการศึกษาครั้งนี้ข้อมูลไม่มีความแตกต่างระหว่างกลุ่ม (ICC=0.2656) ดังนั้นควรพิจารณาเลือกใช้การวิเคราะห์การถดถอยโลจิสติก

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1) ควรมีการจัดกิจกรรมเพื่อรณรงค์และให้สุขภาพ โดยเน้นในเรื่องการออกกำลังกาย การจัดการความเครียด รวมถึงการรับประทานอาหารที่เหมาะสม โดยเฉพาะกลุ่มเสี่ยงที่อายุ 35 ปี ขึ้นไป

2) ควรเน้นจัดกิจกรรมในเรื่องที่เกี่ยวข้องกับโรคเบาหวาน ในกลุ่มผู้ป่วยโรคความดันโลหิตสูง ทั้งในเรื่องของพฤติกรรม การดูแลสุขภาพ เช่น การออกกำลังกาย การจัดการความเครียด และการรับประทานอาหารอย่างเหมาะสม เป็นต้น

3) หากข้อมูลมีแตกต่างระหว่างกลุ่มมาก (ICC \geq 0.5) ควรเลือกใช้สถิติการวิเคราะห์การถดถอยโลจิสติกพหุระดับ แต่หากข้อมูลไม่มีความแตกต่างระหว่างกลุ่มควรเลือกใช้สถิติการวิเคราะห์การถดถอยโลจิสติก

4) หากคำถามการวิจัยมีการพิจารณาผลกระทบระหว่างกลุ่ม (Random effect) ให้พิจารณาเลือกใช้สถิติการวิเคราะห์การถดถอยโลจิสติกพหุระดับ

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1) ควรมีการศึกษาวิเคราะห์โดยใช้แบบจำลองสมการโครงสร้างพหุระดับ (Multilevel structural equation model) เพื่อพิจารณาความสัมพันธ์ระหว่างตัวแปรต้นและการส่งผ่านทั้งทางตรงและทางอ้อมมายังการป่วยเป็นโรคเบาหวาน

2) ควรมีการศึกษาเปรียบเทียบความแม่นยำในวิธีการประมาณค่าที่แตกต่างกัน เช่น การประมาณค่า Maximum likelihood กับ การประมาณค่าแบบเบย์ เพื่อให้ได้วิธีการประมาณค่าที่เหมาะสมต่อการวิเคราะห์พหุระดับในข้อมูลดังกล่าว

3) ควรมีการนำข้อมูลชุดนี้มาวิเคราะห์เหมืองข้อมูล (Data mining) แล้วนำมาเปรียบเทียบกับทั้งสองวิธีการดังกล่าว

กิตติกรรมประกาศ

ขอขอบคุณสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.) สำนัก 3 สำนักสนับสนุนสุขภาพชุมชน ที่ให้ความอนุเคราะห์ข้อมูลในการทำวิจัยครั้งนี้

เอกสารอ้างอิง

1. วชิระ เฟ็งจันทร์. (2558). “เบาหวาน สาเหตุเสียชีวิต อันดับ 3 ในผู้หญิง และอันดับ 8 ในผู้ชาย”. เจาะลึกระบบสุขภาพ. เข้าถึงได้จาก <https://www.hfocus.org/content/2015/11/11245>. [28 สิงหาคม 2562]
2. International diabetes federation. (2018). IDF DIABETES ATLAS. 8th. Brussels: International diabetes federation.
3. สำนักโรคไม่ติดต่อ กระทรวงสาธารณสุข. (2561). ประเด็นสารรณรงค์เบาหวานโลก ปี 2561. นนทบุรี: กระทรวงสาธารณสุข.
4. สำนักโรคเบาหวาน กระทรวงสาธารณสุข. (2559). “โรคเบาหวาน”. สรุปรายงานการเฝ้าระวังโรค ประจำปี 2558. นนทบุรี: กระทรวงสาธารณสุข.

5. Esti Iturralde, Joseph R. Rausch, Jill Weissberg-Benchell, Korey K. Hood. (2019). "Diabetes-Related Emotional Distress Over Time". *Pediatrics*. 143(6): 1-8.
6. N.C. Vaz, A.M. Ferreira, M.S. Kulkarni, F.S. Vaz. (2011). "Prevalence of diabetes mellitus in a rural population of Goa, India". *The National Medical Journal of India*. 24(1): 16-18.
7. Martijn Lappenschaar, Arjen Hommersom, Peter J.F. Lucas, Joep Lagro, Stefan Visscher. (2013). "Multilevel Bayesian networks for the analysis of hierarchical health care data". *Artificial Intelligence in Medicine*. 57(3): 171-183.
8. Ashis Talukder and Zobayer Hossain. (2020). "Prevalence of Diabetes Mellitus and Its Associated Factors in Bangladesh: Application of Twolevel Logistic Regression Model". *Scientific Reports*. 10(10237): 1-7.
9. Randa M. Kutob, Violet Perez Siwik, Mikel Aickin, Cheryl Ritenbaugh. (2014). Families United/Familias Unidas: Family Group Office Visits to Reduce Risk Factors for Type 2 Diabetes. *The Diabetes Educator*. 40(2): 191-201.
10. รติภาคย์ ตามรภาค. (2561). ปัจจัยที่สัมพันธ์กับการเกิดโรคเบาหวานชนิดที่ 2 และประสบการณ์การดูแลตนเองในการควบคุมระดับน้ำตาลในเลือดของผู้สูงอายุชาวไทยภูษาจังหวัดเชียงราย ประเทศไทย. *วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสาธารณสุขศาสตร์ สำนักวิชาวิทยาศาสตร์สุขภาพ มหาวิทยาลัยแม่ฟ้าหลวง*.
11. อกนิษฐ์ ทองจิตร. (2561). การพัฒนาวิธีจำแนกประเภทข้อมูลโดยใช้โครงข่ายประสาทเทียมแบบปรับเหมาะผสมผสานการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค สำหรับการจำแนกประเภทกลุ่มเสี่ยงในการเป็นโรคเบาหวาน. *ดุษฎีนิพนธ์หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา*.
12. กัณทิมา อยู่รวม. (2557). ปัจจัยคัดสรรที่สัมพันธ์กับระดับน้ำตาลเฉลี่ยสะสมในเลือดของข้าราชการตำรวจที่เป็นโรคเบาหวาน ชนิดที่ 2 ที่ไม่สามารถควบคุมระดับน้ำตาลเฉลี่ยสะสมในเลือดได้. *วิทยานิพนธ์ปริญญาพยาบาลศาสตรมหาบัณฑิต สาขาวิชาพยาบาลศาสตรคณะพยาบาลศาสตรจุฬาลงกรณ์มหาวิทยาลัย*.
13. ชลธิดา โยธารินทร์. (2556). ปัจจัยเสี่ยงของโรคเบาหวานชนิดที่ 2 ในประชากรอายุ 40 ปีขึ้นไป ที่ตำบลนามะเขือ อำเภอสหัสขันธ์ จังหวัดกาฬสินธุ์ ประเทศไทย. *วิทยานิพนธ์ปริญญาสาธารณสุขศาสตรมหาบัณฑิต สาขาวิชาสาธารณสุขศาสตร์ วิทยาลัยวิทยาศาสตร์สาธารณสุข จุฬาลงกรณ์มหาวิทยาลัย*.
14. อนัญญา ประดิษฐปรีชา. (2554). ปัจจัยด้านการบริโภคอาหารที่มีผลต่อการเป็นโรคเบาหวาน ชนิดที่ 2 ในประชากร อำเภอนากลาง จังหวัดหนองบัวลำภู. *วิทยานิพนธ์ปริญญาสาธารณสุขศาสตรมหาบัณฑิต สาขาวิชาโภชนศาสตร์เพื่อสุขภาพ บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น*.

15. จารุพร สถุขสิทธิ์สุข. (2553). การพยากรณ์ความเสี่ยงในการเป็นโรคเบาหวานโดยวิธีแบบจำลองสมการโครงสร้าง. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.
16. ชำนาญ สุธงส์นภัย. (2553). การศึกษาและพัฒนาระบบผู้เชี่ยวชาญในการคัดกรองโรคเบาหวาน. วิทยานิพนธ์นี้หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์ เพื่อการศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม.
17. สุดสงวน ชิดโน. (2553). บริโภคนิสัยของประชาชนกลุ่มเสี่ยงต่อการป่วยเป็นโรคเบาหวาน อำเภอนาคู จังหวัดกาฬสินธุ์. วิทยานิพนธ์ปริญญาสาธารณสุขศาสตรมหาบัณฑิต คณะสาธารณสุขศาสตร์ มหาวิทยาลัยบูรพา.
18. วิจิต อินทร์ลำพันธ์. (2549). ภาวะเสี่ยงต่อการเป็นโรคเบาหวานของประชาชนอายุ 40 ปีขึ้นไป จังหวัดสุพรรณบุรี. สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสังคมและการจัดการระบบสุขภาพ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร.
19. Azra Ramezankhani, Kamran Guity, Fereidoun Azizi, and Farzad Hadaegh. (2019). "Sex differences in the association between spousal metabolic risk factors with incidence of type 2 diabetes: a longitudinal study of the Iranian population". *Biology of Sex Differences*. 10(41): 1-8.
20. S. Carlsson, N. Hammar, V. Grill. (2005). Alcohol consumption and type 2 diabetes. *Diabetologia*. 40: 1051-1054.
21. Nick Robinson. (2018). The Disadvantages of Logistic Regression. Retrieve from <https://www.theclassroom.com/disadvantages-statistical-analysis-8471854.html>, [June 28, 2018].
22. Anita R. Kothari and Stephen Birch. (2004). "Multilevel Health Promotion Research: Conceptual and Analytical Considerations". *Canadian Journal of Nursing Research*. 36(1): 56-75.
23. John Neumann. (2016). "Model selection and overfitting". *NATURE METHODS*. 13(9): 703-704.
24. Jindra Reissigova, Zdenek Monhart, Jana Zvarova, Petr Hanzlicek, Hana Grunfeldova, Petr Jansky, Jan Vojacek, Petr Widimsky. (2013). In-hospital Death Prediction by Multilevel Logistic Regression in Patients with Acute Coronary Syndromes. *European Journal of Biomedical Informatics*. 9(1): 11-17.
25. Terry K. Koo, and Mae Y. Li. (2016). "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research". *Journal of Chiropractic Medicine*. 15: 155-163
26. Roxana Alexandrescu, Min-Hua Jen, Alex Bottle, Brian Jarman, and Paul Aylin. (2011). "Logistic Versus Hierarchical Modeling: An Analysis of a Statewide Inpatient Sample". *Journal of American College of Surgeons*. 213(3): 392-401.