

การประเมินการใช้การวิเคราะห์ถดถอย พหุคูณจิสติกในวารสารทางการแพทย์ และสาธารณสุขในประเทศไทยสองวารสาร

กรรณิการ์ ตฤณวุฒิพงษ์*

จิราพร เขียวอยู่**

*สำนักงานป้องกันควบคุมโรคที่ 6 จังหวัดขอนแก่น

**ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาการใช้และการรายงานผลการวิเคราะห์ถดถอยพหุคูณจิสติกในวารสารจดหมายเหตุทางการแพทย์ และวารสารวิชาการสาธารณสุข รวบรวมทุกบทความที่ตีพิมพ์ระหว่างปี พ.ศ. 2545-2550 สืบค้นบทความวิจัยที่ใช้การวิเคราะห์ถดถอยพหุคูณจิสติกด้วยมือ และในฐานข้อมูลอิเล็กทรอนิกส์จากทั้งสองวารสาร คำสำคัญในการสืบค้นได้แก่ “logistic” “regression” และ “multivariable” จากการศึกษาพบว่าบทความที่ใช้การวิเคราะห์ถดถอยพหุคูณจิสติก มี ร้อยละ 6.1 (145/2,370) ในจำนวนนี้ (145) ใช้เพื่อตอบวัตถุประสงค์หลัก ร้อยละ 48.3 (70/145) มีรายงานการกำหนดขนาดตัวอย่างร้อยละ 12.9 (9/70) การทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ ร้อยละ 3.5 (5/145) การทดสอบปฏิกริยาร่วมกันระหว่างตัวแปรอิสระร้อยละ 2.8 (4/145) การทดสอบนัยสำคัญของตัวแบบและตัวแปรอิสระ ร้อยละ 2.1 (3/145) และ 95.2 (138/145) ตามลำดับ และการประเมินสารรูปสมมติ ร้อยละ 4.8 (7/145) ประเด็นที่ไม่มีการรายงานเลยได้แก่ การตรวจสอบความสัมพันธ์เชิงเส้น ซึ่งการวิเคราะห์ที่ไม่ถูกต้องจะส่งผลต่อความน่าเชื่อถือของผลการวิเคราะห์ ซึ่งนำส่งกลับไปในทางที่ผิด และยากต่อการแปลผลได้ และเนื่องจากการวิเคราะห์ถดถอยพหุคูณจิสติกเป็นสถิติที่มีการนำมาใช้มากในงานวิจัยด้านสุขภาพ ดังนั้น นักวิจัย และนักสถิติ ควรให้ความสนใจกับวิธีการทางสถิตินี้ เพื่อให้เกิดความมั่นใจในการใช้หรือการให้คำปรึกษา

คำสำคัญ: การรายงาน, การวิเคราะห์ถดถอยพหุคูณจิสติก

บทนำ

ตัววัดผลลัพธ์หลัก (main outcomes) หรือตัวแปรตามในงานวิจัยทางการแพทย์ และสาธารณสุขจำนวนมากเป็นข้อมูลแจกแจง⁽¹⁾ สถิติที่เหมาะสมในการวิเคราะห์หาความสัมพันธ์ หรือหา

ปัจจัยเสี่ยงระหว่างตัวแปรอิสระหลาย ๆ ตัวแปรกับตัววัดผลลัพธ์หลักประเภทนี้ คือการวิเคราะห์ถดถอยพหุคูณจิสติก⁽²⁾ จากการศึกษาของ Chin⁽³⁾ ซึ่งได้รวบรวมบทความวิจัยจากวารสารด้านระบาดวิทยาพบว่าแนวโน้มการใช้การวิเคราะห์ถดถอยพหุคูณจิสติกเพิ่มขึ้น จาก

เดิมที่ไม่มีรายงานการใช้ใน ค.ศ. 1970 เพิ่มขึ้นเป็นประมาณร้อยละ 3.0 ใน ค.ศ. 1985 และเพิ่มขึ้นเป็นประมาณร้อยละ 16.0 ใน ค.ศ. 2000 นอกจากนี้ยังมีผู้ประเมินการรายงานการวิเคราะห์ถดถอยพหุลอยจิสติกในบทความวิจัยที่ตีพิมพ์ในวารสารวิชาการในต่างประเทศพบว่ามีการรายงานผลการศึกษาที่ไม่ครบถ้วนในหลายประเด็น ได้แก่ ความไม่พอเพียงของอัตราส่วนการเกิดเหตุการณ์ที่สนใจต่อจำนวนตัวแปรอิสระ การทดสอบความสัมพันธ์เชิงเส้น การทดสอบปฏิกริยาร่วมกันระหว่างตัวแปรอิสระ การตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ การประเมินสารรูปสนิทธิ (goodness-of-fit) การรายงานค่า odds ratio และช่วงเชื่อมั่น การคัดเลือกตัวแปรอิสระในการสร้างตัวแบบ การรายงานหน่วยการวัดและรหัสตัวแปรอิสระ และกระบวนการสร้างตัวแบบ⁽⁴⁻⁸⁾ และจากการทบทวนวรรณกรรมของผู้วิจัย ยังไม่พบรายงานวิจัยที่ศึกษาเกี่ยวกับการรายงานผลการวิเคราะห์ถดถอยพหุลอยจิสติกในบทความวิจัยที่ตีพิมพ์ในวารสารด้านการแพทย์และสาธารณสุขในเมืองไทย ผู้วิจัยจึงสนใจศึกษาว่าบทความวิจัยที่ตีพิมพ์ในวารสารทางการแพทย์และสาธารณสุขในประเทศไทยมีการใช้ และการรายงานการวิเคราะห์ถดถอยพหุลอยจิสติกมากน้อยเพียงไร รูปแบบการวิจัยและวัตถุประสงค์ของการใช้ รวมทั้งแนวโน้มการใช้เป็นอย่างไร เพื่อให้ได้ข้อมูลเชิงประจักษ์ในการเสนอแนะแนวทางการใช้และการรายงานผลการวิเคราะห์ถดถอยพหุลอยจิสติกแก่นักวิจัย นักสถิติ และเป็นแนวทางสำหรับผู้อ่านในการประเมินความน่าเชื่อถือของบทความวิจัยได้อย่างถูกต้องและเหมาะสมต่อไป

วิธีการศึกษา

วัตถุประสงค์ ประชากรศึกษาในครั้งนี้คือ บทความวิจัยด้านการแพทย์และสาธารณสุขทุกเรื่อง ซึ่งได้รับการตีพิมพ์เผยแพร่ในสองวารสาร ได้แก่ วารสารจดหมายเหตุทางการแพทย์ และวารสารวิชาการสาธารณสุขระหว่าง พ.ศ. 2545 ถึง พ.ศ. 2550

วิธีการศึกษา เป็นการศึกษาเชิงพรรณนา (descriptive research) ระยะเวลาในการศึกษาระหว่าง เดือนธันวาคม พ.ศ. 2550 ถึง กันยายน พ.ศ. 2551 โดยรวบรวมบทความวิจัยด้านการแพทย์และสาธารณสุขทุกบทความที่ตีพิมพ์ระหว่าง พ.ศ. 2545-2550 สืบค้นบทความวิจัยที่ใช้การวิเคราะห์ถดถอยพหุลอยจิสติก ในกรณีการประเมินความสัมพันธ์ของตัวแปรอิสระแต่ละตัวกับตัวแปรตาม และเป็นเป้าหมายที่ใช้กันมากในงานวิจัยทางการแพทย์และสาธารณสุข โดยการศึกษาครั้งนี้จะไม่ครอบคลุม การวิเคราะห์ถดถอยพหุลอยจิสติกที่มีวัตถุประสงค์เพื่อสร้างสมการทำนายตัวแปรตาม หรือประมาณค่าเพื่อนำไปใช้ในการจำแนกโรค การวินิจฉัยหรือการคัดกรองโดยสืบค้นจากฐานข้อมูลอิเล็กทรอนิกส์และด้วยมือ (manual search) คำสำคัญที่ใช้สืบค้นได้แก่ “logistic” “regression” และ “multivariable” เก็บรวบรวมข้อมูลโดยแบบบันทึกข้อมูล ซึ่งมี 2 ส่วน คือ ลักษณะทั่วไปของบทความวิจัย และประเด็นในการพิจารณาการรายงานการวิเคราะห์ถดถอยพหุลอยจิสติก รายละเอียดดังต่อไปนี้

1. การกำหนดขนาดตัวอย่างในกรณีที่ใช้การวิเคราะห์ถดถอยพหุลอยจิสติกเพื่อตอบคำถามหลัก (primary research question) ซึ่งจะต้องมีการปรับแก้อิทธิพลของความสัมพันธ์ระหว่างตัวแปรอิสระด้วยค่า (variance inflation factor; VIF)⁽⁹⁾ ซึ่งค่า VIF มีค่าเท่ากับ $1/(1-p^2_{1,2,3...p})$ (เมื่อ $p^2_{1,2,3...p}$) คือค่ากำลังสองของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระที่สนใจและตัวแปรอิสระอื่น ๆ จากตัวอย่างที่ได้จากการคำนวณด้วยการวิเคราะห์ถดถอยพหุลอยจิสติกอย่างง่าย (simple logistic regression) โดยเกณฑ์การพิจารณาแบ่งเป็นรายงานครบ รายงาน หรือไม่รายงาน สำหรับบทความวิจัยมีวัตถุประสงค์ในการใช้การวิเคราะห์ถดถอยพหุลอยจิสติกเพื่อตอบวัตถุประสงค์รองพิจารณาความพอเพียงของตัวอย่างในการสร้างตัวแบบจากค่าอัตราส่วนการเกิดเหตุการณ์ที่สนใจต่อตัวแปรอิสระเริ่มต้นซึ่งไม่ควรน้อยกว่า 10⁽¹⁰⁾

2. การคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบเริ่มต้น เกณฑ์การพิจารณาแบ่งเป็น รายงาน หรือไม่รายงาน โดยการรายงานคือ มีการรายงานการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบเริ่มต้น ไม่ว่าจะพิจารณาจากการทบทวนวรรณกรรม และ ความสำคัญทางคลินิก หรือพิจารณา ค่า p-value จากผลการวิเคราะห์คร่าวละตัวแปร

3. การตรวจสอบความสัมพันธ์เชิงเส้น กรณีตัวแปรอิสระเป็นตัวแปรต่อเนื่องหรืออันดับ เกณฑ์การพิจารณาแบ่งเป็น รายงาน หรือไม่รายงาน โดยการรายงานคือ มีการระบุรายละเอียดการตรวจสอบความสัมพันธ์เชิงเส้น เช่น การพล็อตกราฟ การพิจารณาจาก χ^2 for trend

4. การทดสอบปฏิกริยาร่วมกันระหว่างตัวแปรอิสระ (interaction effects) เกณฑ์การพิจารณาแบ่งเป็น รายงาน หรือไม่รายงาน โดยการรายงานคือ มีการกล่าวถึงการพิจารณาจากเนื้อหาของเรื่องที่ศึกษา หรือ จากผลการทดสอบปฏิกริยาร่วมกันระหว่างตัวแปร เช่น จากผล stratified analysis

5. การทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ (collinearity) เกณฑ์การพิจารณาแบ่งเป็น รายงานครบ การรายงาน หรือไม่รายงาน ซึ่งการรายงานครบคือ บทความวิจัยมีการรายงานผลการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ และมีการระบุค่าสถิติในการทดสอบค่าใดค่าหนึ่ง ดังนี้ ค่า VIF หรือ ค่า tolerance หรือ ค่า p^2 การรายงาน คือมีการกล่าวถึงว่ามีการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระแต่ไม่มีรายละเอียดค่าสถิติที่ได้จากการทดสอบ

6. การทดสอบความมีนัยสำคัญทางสถิติของตัวแบบ (model) เกณฑ์การพิจารณาแบ่งเป็น มีการรายงาน หรือไม่รายงาน ซึ่งการรายงานคือมีการรายงานค่าสถิติจากการทดสอบ ได้แก่ ค่า χ^2 หรือ p-value

7. การทดสอบความมีนัยสำคัญทางสถิติของตัวแปรอิสระ เกณฑ์การพิจารณาแบ่งเป็น รายงาน หรือไม่รายงาน โดยการรายงานคือ มีการรายงานข้อใดข้อหนึ่งดังนี้ 1) มีการรายงาน 95%CI ของ β หรือ OR 2)

มีการรายงาน 95%CI ของ β หรือ OR ค่า χ^2 และ รายงานค่า p-value ในส่วนของการไม่รายงาน คือ มีการรายงานค่า χ^2 และค่า p-value หรือรายงานเฉพาะค่า p-value อย่างเดียว

8. การประเมินสารรูปสนิทธิ (goodness-of-fit measures) เกณฑ์การพิจารณาแบ่งเป็น มีการรายงาน หรือไม่รายงาน การรายงานคือ พบรายงานการประเมินความเหมาะสมของตัวแบบ ด้วยวิธีการทดสอบดังนี้ Hosmer-Lemeshow goodness-of-fit test หรือ Pearson chi-square test หรือวิธีอื่น ๆ

9. การรายงานหน่วยการวัดและการให้รหัสของตัวแปรอิสระ (coding of variables) เกณฑ์การพิจารณาแบ่งเป็น การรายงาน หรือไม่รายงาน โดยการรายงานหมายถึง การระบุหรืออธิบายหน่วยการวัด และการให้ค่าตัวเลขของตัวแปรอิสระแต่ละตัวแปร

10. กระบวนการสร้างตัวแบบ (fitting procedure) เกณฑ์การพิจารณาแบ่งเป็น มีการรายงานหรือไม่รายงาน โดยการรายงานคือมีการระบุวิธีการสร้างตัวแบบด้วยวิธีใดวิธีหนึ่งดังนี้ วิธี forward backward หรือ step-wise

การวิเคราะห์ข้อมูล บรรยายรายละเอียดของการใช้ การรายงานการวิเคราะห์ถดถอยพหุคูณในแต่ละประเด็นพิจารณาด้วยสถิติพรรณนา นำเสนอด้วยค่าความถี่ ร้อยละ และใช้สถิติอนุมานวิเคราะห์แนวโน้ม การใช้การวิเคราะห์ถดถอยพหุคูณด้วยสถิติ chi-square for trend และประเมินความเหมาะสมของสมการ (goodness-of-fit) กำหนดระดับนัยสำคัญในการทดสอบเท่ากับ 0.05 วิเคราะห์ข้อมูลด้วยโปรแกรม STATA version 8.2

ผลการศึกษา

บทความวิจัยที่ได้รับการตีพิมพ์เผยแพร่ในวารสารด้านการแพทย์และสาธารณสุขของเมืองไทยสองวารสารระหว่าง พ.ศ. 2545 ถึง พ.ศ. 2550 จำนวนทั้งหมด 2,370 บทความ ตีพิมพ์ในวารสารจดหมายเหตุทางการแพทย์

ร้อยละ 81.2 (1,924/2,370) และวารสารวิชาการสาธารณสุข ร้อยละ 18.8 (446/2,370) มีบทความวิจัยที่ใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติกในการวิเคราะห์ข้อมูลทั้งหมด 145 บทความ คิดเป็นร้อยละ 6.1 โดยตีพิมพ์ในวารสารจดหมายเหตุทางการแพทย์ และวารสารวิชาการสาธารณสุข ร้อยละ 81.4 (118/145) และ 18.6 (27/145) ตามลำดับ

แนวโน้มการใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติกเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ โดยในภาพรวมเพิ่มจากร้อยละ 11.0 ใน พ.ศ. 2545 เป็นร้อยละ 21.4 ใน พ.ศ. 2550 (p-value 0.0017) และในวารสารจดหมายเหตุทางการแพทย์ เพิ่มจากร้อยละ 7.6 ใน พ.ศ. 2545 เป็นร้อยละ 22.9 ใน พ.ศ. 2550 (p-value 0.0002) รายละเอียดดังตารางที่ 1

ผลการพิจารณาการรายงานการวิเคราะห์ถดถอยพหุคูณโลจิสติกในแต่ละประเด็นพิจารณา

การพิจารณาการรายงานการวิเคราะห์ถดถอยพหุคูณโลจิสติกทั้งหมด 10 ประเด็น พบว่า ประเด็นการรายงานที่พบมากที่สุดคือ การทดสอบความมีนัยสำคัญ

ทางสถิติของตัวแปรอิสระแต่ละตัวแปร ร้อยละ 95.2 รองลงมาคือกรายงานการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบเริ่มต้น ร้อยละ 92.4 และมีประเด็นการพิจารณาที่ไม่มีบทความใดรายงานเลยได้แก่ การตรวจสอบความสัมพันธ์เชิงเส้น กรณีที่มีตัวแปรอิสระเป็นตัวแปรต่อเนื่องหรืออันดับ (ตารางที่ 2)

วิจารณ์

จากการศึกษา พบว่าแนวโน้มการใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติกเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติโดยในภาพรวมเพิ่มจากร้อยละ 11.0 ใน พ.ศ. 2545 เป็นร้อยละ 21.4 ใน พ.ศ. 2550 (p-value 0.0017) และในวารสารจดหมายเหตุทางการแพทย์ เพิ่มจากร้อยละ 7.6 ใน พ.ศ. 2545 เป็นร้อยละ 22.9 ใน พ.ศ. 2550 (p-value 0.0002) ซึ่งสอดคล้องกับการศึกษาของ Chin⁽³⁾ ซึ่งได้รวบรวมบทความวิจัยด้านระบาดวิทยาระหว่าง ค.ศ. 1970 ถึง ค.ศ. 2000 จำนวน 4 วารสาร พบว่าแนวโน้มการใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติกเพิ่มขึ้นจากเดิมที่ไม่มีรายงานการใช้ใน ค.ศ. 1970 เพิ่มขึ้นเป็นร้อยละ 16.0 ใน ค.ศ. 2000 และ Bagley⁽⁵⁾ ได้สรุปว่า

ตารางที่ 1 แนวโน้มการใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติก ระหว่าง พ.ศ. 2545 ถึง พ.ศ. 2550 (n = 2,370 บทความ)

พ.ศ.	จดหมายเหตุทางการแพทย์		วารสารวิชาการสาธารณสุข		รวมจำนวนบทความ	
	x/n	ร้อยละ	x/n	ร้อยละ	x/n	ร้อยละ
2545	9/338	7.6	7/80	25.9	16/418	11.0
2546	16/225	13.6	3/85	11.1	19/310	13.1
2547	10/283	8.5	5/86	18.5	15/369	10.3
2548	29/456	24.6	6/53	22.2	35/509	24.1
2549	27/348	22.6	2/75	7.4	29/423	20.0
2550	27/274	22.9	4/67	14.8	31/341	21.4
รวม	118/1,924	6.1	27/446	6.1	145/2,370	6.1
p-value ^a		0.0002		0.5912		0.0017
p-value ^b		0.2424		0.2151		0.5028

หมายเหตุ: x = บทความที่ใช้ Logistic Regression, n = บทความทั้งหมด

a = Analyse Data by Chi-square (χ^2) for Trend, b = Goodness of Fit by Chi-square (χ^2) for Departure

การประเมินการใช้การวิเคราะห์ถดถอยพหุคูณในวารสารทางการแพทย์และสาธารณสุขในประเทศไทยสองวารสาร

ตารางที่ 2 ประเด็นการพิจารณาการรายงานการวิเคราะห์ถดถอยพหุคูณ จําแนกรายวารสารที่ศึกษา

ประเด็นพิจารณา	จดหมายเหตุทางการแพทย์		วารสารวิชาการสาธารณสุข		รวม	
	(n = 118 บทความ)		(n = 27 บทความ)		(n = 145 บทความ)	
	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
การกำหนดขนาดตัวอย่าง พิจารณากรณีตอบวัตถุประสงค์หลัก (n=70)						
รายงานครบ	-	-	1	7.69	1	1.43
รายงาน	7	12.28	1	7.69	8	11.43
ไม่พบรายงาน	50	87.82	11	84.62	61	87.14
การคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบเริ่มต้น						
รายงาน	110	93.2	24	88.9	134	92.4
ไม่รายงาน	8	6.8	3	11.1	11	7.6
การทดสอบปฏิกริยาร่วมกันระหว่างตัวแปรอิสระ						
รายงาน	4	3.4	-	-	4	2.8
ไม่รายงาน	114	96.6	27	100.0	141	97.2
การทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ						
รายงาน	5	4.2	-	-	5	3.5
ไม่รายงาน	113	95.8	27	100.0	140	96.5
การทดสอบความมีนัยสำคัญทางสถิติของตัวแบบ						
รายงาน	3	2.5	-	-	3	2.1
ไม่รายงาน	115	97.5	27	100.0	142	97.9
การทดสอบความมีนัยสำคัญทางสถิติตัวแปรอิสระแต่ละตัว						
รายงาน	111	94.1	27	100.0	138	95.2
ไม่รายงาน	7	5.9	-	-	7	4.8
การประเมินสารรูปสถิติ						
รายงาน	6	5.1	1	3.7	7	4.8
ไม่รายงาน	112	94.9	26	96.3	138	95.2
การรายงานระดับการวัดและการให้รหัสของตัวแปรอิสระ						
รายงาน	108	91.5	24	88.9	132	91.0
ไม่รายงาน	10	8.5	3	11.1	13	9.0
กระบวนการสร้างตัวแบบ						
รายงาน	24	20.3	3	11.1	27	18.6
ไม่รายงาน	94	79.7	24	88.9	118	81.4

เหตุผลในการเลือกใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติก เนื่องจากตัววัดผลลัพธ์หลักเป็นตัวแปรเจนนับ และงานวิจัยจำนวนมากทางด้านการแพทย์และสาธารณสุขต้องการหาปัจจัยเสี่ยงหรือปัจจัยที่มีความสัมพันธ์ของตัวแปรอิสระหลายตัวกับตัวแปรตาม และอีกเหตุผลหนึ่งเนื่องจากได้มีการพัฒนาโปรแกรมสำเร็จรูปทางสถิติและได้นำมาใช้อย่างแพร่หลาย ทำให้วิเคราะห์ข้อมูลได้และผลการวิเคราะห์ที่ได้มีความแม่นยำ

ผลจากการพิจารณาการรายงานการวิเคราะห์ถดถอยพหุคูณโลจิสติกในแต่ละประเด็น อภิปรายผลได้ดังนี้

1. การคำนวณขนาดตัวอย่าง ในภาพรวมมีบทความวิจัยที่มีวัตถุประสงค์ในการใช้การวิเคราะห์ถดถอยพหุคูณโลจิสติกเพื่อตอบวัตถุประสงค์หลัก จำนวน 70 บทความ ในจำนวนนี้มีการรายงานการกำหนดขนาดตัวอย่างสำหรับการวิเคราะห์ถดถอยพหุคูณโลจิสติก ร้อยละ 12.9 (9/70) เมื่อจำแนกตามรายวารสาร พบว่า วารสารจดหมายเหตุทางการแพทย์มีการรายงาน ร้อยละ 12.3 (7/57) และวารสารวิชาการสาธารณสุขมีการรายงาน ร้อยละ 15.4 (2/13) ในส่วนบทความวิจัยที่มีวัตถุประสงค์เพื่อตอบวัตถุประสงค์รอง จำนวน 75 บทความ ในจำนวนนี้สามารถประเมินจำนวนการเกิดเหตุการณ์ที่สนใจต่อตัวแปรอิสระเริ่มต้นได้ 61 บทความ เมื่อคำนวณอัตราส่วนการเกิดเหตุการณ์ที่สนใจต่อตัวแปรอิสระเริ่มต้นโดยพิจารณาจากทั้งสองวารสารรวมกัน พบว่า มีบทความที่มีค่าอัตราส่วนการเกิดเหตุการณ์ที่สนใจต่อตัวแปรอิสระเริ่มต้นน้อยกว่า 10.0 ร้อยละ 45.9 (28/61) เมื่อจำแนกตามรายวารสารพบว่า วารสารจดหมายเหตุทางการแพทย์ และวารสารวิชาการสาธารณสุขมีอัตราส่วนการเกิดเหตุการณ์ที่สนใจต่อตัวแปรอิสระเริ่มต้นน้อยกว่า 10.0 ร้อยละ 44.9 (22/49) และ 50.0 (6/12) ตามลำดับ ซึ่งถือว่ามีขนาดตัวอย่างที่ไม่เหมาะสมและจะส่งผลต่อการประมาณค่าสัมประสิทธิ์ถดถอยซึ่งอาจจะ overestimate ช่วงเชื่อมั่นของค่า OR ไม่กระชับ^(4-6,8) และ Concato⁽⁶⁾ ได้สรุปว่าผลจากความไม่พอเพียงดังกล่าวจะทำให้การทดสอบความมีนัย-

สำคัญอาจไม่ถูกต้อง ดังนั้น จึงควรวางแผน หรือคำนวณขนาดตัวอย่างก่อนการเก็บรวบรวมข้อมูลสำหรับบทความวิจัยที่ตอบวัตถุประสงค์รองก็ควรพิจารณาความพอเพียงจากค่าอัตราส่วนการเกิดเหตุการณ์ที่สนใจต่อตัวแปรอิสระเริ่มต้นซึ่งไม่ควรน้อยกว่า 10.0⁽¹⁰⁾

2. การคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบเริ่มต้น ในภาพรวมมีการรายงาน ร้อยละ 92.4 จำแนกตามการพิจารณาจากงานวิจัยที่ผ่านมาหรือความสำคัญทางคลินิก ร้อยละ 100.0 และพิจารณาการนำเข้าสู่ตัวแปรอิสระจากผลการวิเคราะห์คร่าวละตัวแปร ร้อยละ 47.0 โดยค่า p-value ที่ใช้ในการพิจารณานำเข้าสู่ตัวแบบเริ่มต้นที่พบมากที่สุด คือกำหนดค่า p-value น้อยกว่า 0.05 (84.1%) ซึ่งในกรณีที่ผู้วิจัยกำหนดระดับนัยสำคัญที่ $\alpha < 0.05$ การสร้างตัวแบบอาจจะเกิดความไม่เหมาะสมเนื่องจากตัวแปรบางตัวแปรที่มีความสำคัญแต่ค่า p-value สูงกว่าระดับนัยสำคัญที่กำหนด จะทำให้ตัวแปรนั้นไม่ถูกนำเข้าสู่ตัวแบบได้ หรือในกรณีที่กำหนดระดับนัยสำคัญไว้สูงเกินไปก็จะทำให้ได้ตัวแปรที่ไม่มีความสำคัญเข้ามาในตัวแบบ^(2,11) ดังนั้น ควรกำหนดระดับนัยสำคัญในการพิจารณานำตัวแปรอิสระเข้ามาสร้างตัวแบบให้มีค่าอยู่ระหว่าง 0.1-0.2 หรือสูงสุดไม่เกิน 0.25⁽²⁾ และ Lang⁽¹²⁾ กล่าวว่า เกณฑ์ในการกำหนดการนำเข้าสู่ตัวแปรอิสระจากการวิเคราะห์คร่าวละตัวแปรควรมีค่า p-value น้อยกว่า 0.1 จึงจะถูกพิจารณานำเข้าในการสร้างตัวแบบ

3. การตรวจสอบความสัมพันธ์เชิงเส้น พบว่าบทความวิจัยที่มีตัวแปรอิสระเป็นตัวแปรต่อเนื่องหรืออันดับ ไม่มีการรายงานการตรวจสอบความสัมพันธ์เชิงเส้น สอดคล้องกับการศึกษาของ Bagley⁽⁵⁾ ที่ศึกษาใน Journal of Clinical Epidemiology ตีพิมพ์ ค.ศ.1985 และ ค.ศ.1999 จำนวนทั้งหมด 21 บทความ ผลการศึกษาไม่พบการรายงานการทดสอบความสัมพันธ์เชิงเส้นทุกบทความวิจัย

4. การทดสอบปฏิกริยาร่วมกันระหว่างตัวแปรอิสระ ในภาพรวมมีการรายงานเพียง ร้อยละ 2.8 ซึ่ง

ผู้ประเมินไม่แน่ใจว่า ข้อมูลนี้ผู้วิจัยมีการทดสอบ ปฏิกริยาร่วมกันระหว่างตัวแปรอิสระแต่ผลการตรวจสอบไม่มีการเกิดปฏิกริยาร่วมกันระหว่างตัวแปรอิสระ จึงไม่นำมาเสนอไว้ในรายงานวิจัย หรืออีกกรณีหนึ่งคือ ผู้วิจัยไม่ได้มีการทดสอบตั้งแต่ต้นก่อนการสร้างตัวแบบ จึงไม่มีการนำเสนอหรือรายงานในบทความ ซึ่งสอดคล้องกับการศึกษาของ Bagley⁽⁵⁾ และ Ottenbacher⁽⁴⁾ ที่รายงานว่าผู้ประเมินไม่แน่ใจว่าการไม่รายงานการทดสอบเกิดจากสาเหตุใดดังที่กล่าวมา

5. การทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ ในภาพรวมมีการรายงาน ร้อยละ 3.5 และทุกบทความตีพิมพ์ในวารสารจดหมายเหตุทางการแพทย์ โดยการรายงานเพียงแต่ระบุว่ามีการพิจารณาตัวแปรที่มีความสัมพันธ์กันจะนำเข้าตัวแปรตัวใดตัวหนึ่งในการสร้างตัวแบบ แต่ไม่ได้ระบุรายละเอียดการพิจารณา ซึ่งการที่ตัวแปรอิสระมีความสัมพันธ์ระหว่างกันสูงแสดงว่าเกิดปัญหา multicollinearity จะส่งผลให้ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์ถดถอยมีแนวโน้มสูงขึ้น (inflated) และบางครั้งการประมาณค่าสัมประสิทธิ์ถดถอยจะไม่น่าเชื่อถือ⁽¹³⁾ สามารถพิจารณาจากค่า VIF หรือ ค่า tolerance ถ้าค่า VIF มากกว่า 10.0 หรือค่า tolerance น้อยกว่า 0.1 แสดงว่าตัวแปรอิสระที่สนใจมีความสัมพันธ์กับตัวแปรอิสระอื่น ๆ ในตัวแบบสูงและรุนแรง Wetherill⁽¹⁴⁾

6. การทดสอบความมีนัยสำคัญทางสถิติของตัวแบบ เป็นการเปรียบเทียบระหว่างตัวแบบที่ไม่มีตัวแปรอิสระที่ต้องการทดสอบ และตัวแบบที่มีตัวแปรอิสระที่ต้องการทดสอบ สถิติที่ใช้ คือ Likelihood ratio test (G) จากการศึกษาครั้งนี้ พบว่า ในภาพรวมมีการรายงานการทดสอบความมีนัยสำคัญทางสถิติของตัวแบบ ร้อยละ 2.1 โดยในจำนวนนี้มีการรายงานการทดสอบด้วยค่าไคสแควร์ (χ^2) และ p-value

7. การทดสอบความมีนัยสำคัญทางสถิติของตัวแปรอิสระแต่ละตัว ได้แก่ การทดสอบโดย Likelihood ratio test (G) และ Wald test จากการศึกษา

พบว่า มีการรายงานในภาพรวม ร้อยละ 95.2 โดยการรายงานที่พบมากที่สุดคือการรายงานค่า 95%CI ของ OR ร้อยละ 94.9 (130/138) ซึ่งในการรายงานผลการทดสอบนัยสำคัญทางสถิติของตัวแปรอิสระแต่ละตัวในตัวแบบสุดท้ายควรจะมีการนำเสนอค่า p-value ค่า OR และ 95%CI⁽⁵⁾

8. การประเมินสารรูปสถิติ เป็นการประเมินความเหมาะสมของตัวแบบ สถิติที่ใช้ในการประเมินมีหลายตัว ได้แก่ Pearson Chi-square (χ^2) Hosmer-Lemeshow goodness-of-fit test หรือ Likelihood-ratio test (G2) จากการศึกษาครั้งนี้ พบว่า ในภาพรวมบทความวิจัยที่ไม่รายงานร้อยละ 95.2 โดยมีบทความที่รายงานจำนวน 7 บทความ สถิติที่ใช้ทดสอบคือ Likelihood ratio test จำนวน 4 บทความ และ Hosmer-Lemeshow goodness-of-fit test จำนวน 2 บทความ ซึ่งบทความวิจัยที่รายงานถือว่ามีสัดส่วนที่น้อยมาก สอดคล้องกับการศึกษาของ Khan⁽⁷⁾ ที่พบว่า มีบทความวิจัยที่ไม่รายงานการประเมินสารรูปสถิติสูงถึงร้อยละ 93.2

9. การรายงานหน่วยการวัดและการให้รหัสของตัวแปรอิสระ ในภาพรวมมีการรายงาน ร้อยละ 91.0 แต่ไม่มีบทความใดที่รายงานการให้รหัสตัวแปรอิสระได้อย่างสมบูรณ์คือรายงานทั้งหน่วยการวัดและการให้รหัสของตัวแปรอิสระ ซึ่งสอดคล้องกับการศึกษาของ Bagley⁽⁵⁾ ที่พบว่าไม่มีบทความวิจัยที่กล่าวถึงหรือรายงานการให้รหัสตัวแปรได้สมบูรณ์ โดยมีการรายงานเพียงว่าตัวแปรเพศ มีเพศชายและหญิงเท่านั้น ซึ่งการระบุหน่วยและรหัสของตัวแปรอิสระจะสำคัญมากหากมีตัวแปรอิสระที่เป็น interaction term ในตัวแบบ⁽⁵⁾ โดย Lang⁽¹²⁾ กล่าวว่า ในการให้รหัสของตัวแปรอิสระไม่เพียงแต่จะระบุตัวแปรเพศเป็นชายหรือหญิงเท่านั้น แต่ควรจะต้องระบุเป็นตัวเลขที่ให้ค่า เช่น ชาย เท่ากับ 1 หญิง เท่ากับ 0

10. การสร้างตัวแบบ ในการพิจารณาเลือกตัวแปรอิสระเข้ามาในตัวแบบมีด้วยกันหลายวิธี ได้แก่ วิธี

forward backward และ stepwise ซึ่งวิธีการต่าง ๆ จะทำให้ได้ตัวแบบที่แตกต่างกันถึงแม้ว่าจะเป็นข้อมูลชุดเดียวกันก็ตาม ซึ่ง Kleinbaum⁽¹⁾ ได้เสนอว่า ในการพิจารณาตัวแปรกวนในการสร้างตัวแบบเพื่อประเมินความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรผลลัพธ์หลัก ไม่ควรที่จะสร้างโดยใช้คำสั่งให้โปรแกรมสำเร็จรูปทางสถิติสร้างตัวแบบเอง เช่น วิธี Forward หรือ Backward ควรจะพิจารณาตัวแปรกวนโดยคำนึงถึงเนื้อหา (content) ของเรื่องที่ศึกษาโดยผู้วิจัยเอง

สรุป

ในภาพรวมบทความวิจัยมีแนวโน้มการใช้การวิเคราะห์ถดถอยพหุคูณจิสติกในการวิเคราะห์ข้อมูลเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ $p\text{-value} < 0.05$ ประเด็นการพิจารณาที่ไม่มีบทความวิจัยใดรายงานเลยได้แก่ การตรวจสอบความสัมพันธ์เชิงเส้น ในกรณีที่มีตัวแปรอิสระเป็นตัวแปรต่อเนื่องหรือตัวแปรอันดับ ในส่วนของประเด็นอื่น ๆ ที่มีการรายงานน้อย ได้แก่ การกำหนดขนาดตัวอย่าง การทดสอบปฏิกริยาร่วมกันระหว่างตัวแปรอิสระ การทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ และการทดสอบความมีนัยสำคัญของตัวแบบ ซึ่งเป็นประเด็นสำคัญที่ควรพิจารณาเมื่อวิเคราะห์ข้อมูลและควรรายงานผลการพิจารณาในประเด็นเหล่านี้ด้วย และเนื่องจากการวิเคราะห์ถดถอยพหุคูณจิสติกเป็นสถิติที่มีการนำมาใช้มากในงานวิจัยด้านสุขภาพ ในการวิเคราะห์ข้อมูลมีความจำเป็นอย่างยิ่งที่นักวิจัยจะต้องตระหนักถึงความน่าเชื่อถือของผลการศึกษาดังนั้น ในกระบวนการดำเนินการเกี่ยวกับข้อมูล ตั้งแต่การวางแผนก่อนการดำเนินการจนถึงการรายงานผลการศึกษา จะต้องพิจารณาถึงประเด็นที่เกี่ยวข้องกับการใช้และการรายงานผลการวิเคราะห์ถดถอยพหุคูณจิสติกทั้ง 10 ประเด็นที่กล่าวมาข้างต้น ซึ่งผู้เชี่ยวชาญได้เสนอแนะให้มีการพิจารณาในการใช้และการรายงานการวิเคราะห์ถดถอยพหุคูณจิสติก จากการศึกษาครั้งนี้จะเห็นได้ว่าบทความวิจัยที่ตีพิมพ์ในวารสารด้านการ

แพทย์และสาธารณสุขทั้งสองวารสาร ยังมีการรายงานในประเด็นที่พิจารณาไม่ครบถ้วน ซึ่งการไม่รายงานดังกล่าวผู้วิจัยอาจจะมีการทดสอบหรือดำเนินการในแต่ละประเด็นแต่ไม่ได้นำผลมานำเสนอในบทความวิจัยก็เป็นได้ แต่เพื่อให้ผู้อ่านได้ใช้เป็นแนวทางในการประเมินความน่าเชื่อถือของผลการศึกษานักวิจัยและนักสถิติ ควรให้ความสนใจกับการใช้ และประเด็นการรายงานของวิธีการทางสถิตินี้เพื่อให้เกิดความมั่นใจในการใช้ หรือการให้คำปรึกษาต่อไป

กิตติกรรมประกาศ

ขอขอบคุณ รองศาสตราจารย์ ดร. บัณฑิต ถิ่นคำพร รองศาสตราจารย์ ดร. นพ. สมพงษ์ ศรีแสนปวง ผู้ช่วยศาสตราจารย์ ดร.จรมนิต แก้วกั้งวาล และผู้ช่วยศาสตราจารย์ นิคม ถนอมเสียง ที่กรุณาตรวจสอบเครื่องมือ และให้ข้อเสนอแนะให้งานวิจัยนี้มีความสมบูรณ์ยิ่งขึ้น

เอกสารอ้างอิง

1. Kleinbaum GD. Logistic regression: a self-learning text. New York: Springer-Verlag; 1994.
2. Hosmer WD, Lemeshow S. Applied logistic regression. 2nd ed. The United States of America: John Wiley & Sons; 2000.
3. Chin S. The rise and fall of logistic regression. Aust Epidemiol 2001; 8:7-10.
4. Ottechinacher JK, Ottenbacher RH, Tooth L, Ostir VG. A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. J Clin Epidemiol 2004; 57:1147-52.
5. Bagley CS, White H, Golomb AB. Logistic regression in the medical literature: standard for use and reporting, with particular attention to one medical domain. J Clin Epidemiol 2001; 54:979-85.
6. Concato J, Feinstein RA, Holford RT. The risk of determining risk with multivariable models. Ann Intern Med 1993; 118:201-10.
7. Khan SK, Chien FWP, Dwarakanath SL. Logistic regression models in obstetrics and gynecology literature. Obstet Gynecol 1999; 93(6):1014-20.

8. Moss M, Wellman DA, Cotsonis AG. An appraisal of multivariable logistic models in the pulmonary and critical care literature. CHEST 2003; 123:923-8.
9. Hsieh YF, Bloch AD, Larsen DM. A simple method of sample size calculation for linear and logistic regression. Statistics in Medicine 1998; 17:1623-34.
10. Peduzzi P, Concato J, Kemper E, Holford RT, Feinstein RA. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49(12):1373-9.
11. Sun WG, Shook LT, Kay LG. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996; 49(8): 907-16.
12. Lang T. Documenting research in scientific articles: guidelines for authors: 3. Reporting multivariate analysis. CHEST 2007; 131: 628-32.
13. UCLA Academic Technology Services. Lesson 3 logistic regression diagnostics. [cited 2007 Oct 31]; Available from: URL: <http://www.ats.ucla.edu/stat/Stata/webbooks/logistic/chapter3/stalog3.htm>
14. Wetherill G. Regression analysis with applications. London: Chapman and Hall; 1986.

Abstract An Appraisal of Logistic Regression Models Reported in Published Articles in Two Thai Health Journals

Kannika Tinnawutipong*, Jiraporn Khiewyoo**

*Office of Disease Prevention and Control 6th Khon Kaen, **Department of Biostatistics and Demography, Faculty of Public Health, Khon Kaen University

Journal of Health Science 2009; 18:665-73.

This study was to determine the using and reporting of multiple logistic regression analysis in Journal of the Medical Association of Thailand and Journal of Health Science from 2002 to 2007. All original papers were searched by manual and electronic data based for the use of multiple logistic regression, matching keywords as follows; "logistic" "regression" and "multivariable". Multiple logistic regression was used 6.1 percent (145/2,370), in these papers, reporting of logistic regression analysis as follows; used for the primary research question 48.3 percent (70/145) and sample size calculation was reported 12.9 percent (9/70), assessment of multicollinearity 3.5 percent (5/145), testing for interactions 2.8 percent (4/145), test of model and independent variables significantly 2.1 (3/145) and 95.2 percent (138/145) respectively, and goodness-of-fit measures 4.8 percent (7/145). In addition, no studies reported test of conformity of variables to a linear gradient. These problems would make the reported results potentially inaccurate, misleading, or difficult to interpret. Multiple logistic regression has become a standard statistical method in health and medical researches. It is recommended that researchers and statisticians pay greater attention to guidelines concerning the use and reporting of multiple logistic regression.

Key words: reporting, multiple logistic regression analysis