

*Review Article***Principle of Meta-analysis of Clinical Trials**Nalinee Poolsup

*Department of Pharmacy, Faculty of Pharmacy, Silpakorn University, Sanamchandra Palace Campus, Nakhon Pathom 73000.
Email: nalinee@email.pharm.su.ac.th*

Abstract

A meta-analysis is a type of systematic review that employs statistical methods to synthesize the evidence from several primary studies. It has been increasingly applied to combine results from randomized controlled trials of health interventions. Several statistical methods are available for combining the data in meta-analysis. If conducted properly, meta-analysis can provide useful information for clinical decision making. Despite its value, meta-analysis has been subjected to criticism. The statistical basis and the potentials and problems of meta-analysis are considered in this review.

Key words

Meta-analysis, Systematic review, Evidence-based pharmacotherapy.

นิพนธ์ปฏิบัติ

การสังเคราะห์ข้อมูลโดยวิธี meta-analysis†

นลินี พูลทรัพย์

ภาควิชาเภสัชกรรม คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ อ. เมือง จ. นครปฐม 73000

E-mail: nalinee@email.pharm.su.ac.th

บทคัดย่อ

Meta-analysis เป็นการทบทวนวรรณกรรมอย่างเป็นระบบที่มีการใช้วิธีทางสถิติที่เหมาะสมในการรวมผลที่ได้จากงานวิจัยตั้งแต่สองเรื่องขึ้นไปเข้าด้วยกัน ปัจจุบันได้มีการนำเทคนิคนี้มาใช้กันอย่างแพร่หลายมากขึ้นในการสังเคราะห์ผลที่ได้จากการศึกษาเปรียบเทียบวิธีต่างๆที่เกี่ยวข้องกับการดูแลสุขภาพและมีบทความประเภทยาตีพิมพ์เผยแพร่มากขึ้น ถึงแม้ว่า meta-analysis จะมีข้อดีซึ่งเป็นที่ยอมรับกันอย่างกว้างขวางแต่ก็ยังมีข้อจำกัดอยู่หลายประการ บทความนี้กล่าวถึงพารามิเตอร์ทางสถิติที่ใช้แสดงผลการเปรียบเทียบระหว่างกลุ่มที่ศึกษา สถิติที่ใช้ในการรวมผลการวิจัย การทดสอบความคล้ายคลึงกันของผลการวิจัยและสาเหตุของความแตกต่างกันของผลการวิจัย รวมทั้งปัญหาที่พบใน meta-analysis

กุญแจคำ

การวิเคราะห์เมตา การทบทวนวรรณกรรมอย่างเป็นระบบ การเลือกใช้ยานหลักฐานทางวิชาการ

วารสารทางการแพทย์ในปัจจุบันมีการตีพิมพ์บทความจำนวนมากที่เป็นการรวบรวมและสังเคราะห์ผลการวิจัยที่มีการศึกษามาก่อนเข้าไว้ด้วยกัน เรียกการวิจัยแบบนี้ว่า "meta-analysis" ซึ่งคำนี้ถูกนำมาใช้เป็นครั้งแรกโดย Glass (1) ซึ่งเป็นนักวิจัยทางการศึกษา โดยได้ให้ความหมายไว้ว่า เป็นการใช่วิธีทางสถิติในการวิเคราะห์ผลการวิเคราะห์ที่ได้จากงานวิจัย (analysis of analyses) เพื่อประมวลผลที่ได้เข้าด้วยกัน หลังจากนั้นได้มีความสนใจในการสังเคราะห์งานวิจัยในเชิงปริมาณ (quantitative synthesis) มากขึ้นในหมู่นักวิจัยเชิงสังคมศาสตร์ โดยเฉพาะทางจิตวิทยา แล้วจึงขยายมายังการวิจัยทางการแพทย์ สาธารณสุข และการบริการด้านสุขภาพ (2) ความ-

หมายอื่นของ meta-analysis ที่ได้มีการให้ความหมายกันไว้ ได้แก่ การรวมกันอย่างเป็นระบบ (systematic quantitative pooling) โดยใช้วิธีทางสถิติที่เหมาะสมในการสังเคราะห์ผลที่ได้จากงานวิจัยที่ศึกษาปัญหาวิจัยเดียวกัน (3) หรืออาจหมายถึงการทบทวนวรรณกรรมอย่างเป็นระบบ (systematic review) ที่มีการใช้วิธีทางสถิติในการรวมผลที่ได้จากงานวิจัยตั้งแต่สองเรื่องขึ้นไปเข้าด้วยกัน (4) นอกจากนี้ ยังมีคำอื่นๆที่ใช้ในความหมายเดียวกันอีก ได้แก่ quantitative systematic review (4), quantitative overview (5), data pooling, pooled analysis, integrative research review, data synthesis, quantitative synthesis, และ combining studies (6) meta-analysis จึงเป็นส่วนหนึ่งใน systematic review ซึ่งเข้ามามีบทบาทที่สำคัญมากขึ้น โดยเฉพาะในปัจจุบันที่จำเป็น

†ส่วนหนึ่งนำเสนอในการประชุมวิชาการประจำปี 2544 สมาคมเภสัชกรรมโรงพยาบาล (ประเทศไทย)

ต้องอาศัยหลักฐานทางวิชาการที่ดีที่สุดที่มีอยู่ในขณะนั้นมาช่วยประกอบในการตัดสินใจในการดูแลผู้ป่วย (evidence-based practice) ไม่ว่าจะอยู่ในระดับรายบุคคลหรือในการกำหนดนโยบายที่เกี่ยวข้องกับการดูแลสุขภาพ สำหรับในประเทศไทย การวิจัยในแขนงนี้ยังจำกัดอยู่ในแวดวงการศึกษา โดยมีการแปลศัพท์ "meta-analysis" เป็นภาษาไทยไว้แตกต่างกันไป ได้แก่ การวิเคราะห์เมตต้า การวิเคราะห์เมตต้า การวิเคราะห์อภิมาน การวิเคราะห์รวมผล และการอภิเคราะห์ (7) บทความนี้จะขอใช้คำว่า "meta-analysis"

วัตถุประสงค์ของ meta-analysis

Meta-analysis หรือ systematic quantitative review เป็นวิธีที่มีแบบแผนขั้นตอนชัดเจนในการรวบรวมและสังเคราะห์ข้อมูลหลักฐานที่ได้จากงานวิจัยที่ต้องการตอบปัญหาเดียวกันเข้าด้วยกัน ทำให้จำนวนตัวอย่างที่ศึกษามีมากขึ้น จึงช่วยเพิ่มความสามารภในการแสดงความแตกต่างทางสถิติ (statistical power) ระหว่างกลุ่มที่เปรียบเทียบกัน และเพิ่มความแม่นยำ (precision) ของผลที่ได้ รวมทั้งเป็นวิธีที่ได้มาซึ่งข้อสรุปสุดท้ายในกรณีที่มีความแตกต่างหรือความขัดแย้งกันระหว่างผลการวิจัยเหล่านั้น นอกจากนี้ยังช่วยตอบคำถามอื่นที่ไม่ได้ตั้งขึ้นหรือไม่ได้เป็นจุดสนใจเมื่อแรกเริ่มทำการวิจัย รวมทั้งคำถามที่จำเพาะต่อผู้ป่วยกลุ่มย่อยๆ ซึ่งไม่สามารถหาคำตอบได้ในงานวิจัยแต่ละเรื่องอันเนื่องมาจากกลุ่มตัวอย่างที่ศึกษามีขนาดเล็กเกินไป โดยเฉพาะอย่างยิ่งในการวิจัยแบบ randomized controlled trial (8) นอกจากนี้ผลที่ได้จาก meta-analysis อาจชี้ให้เห็นถึงปัญหาหรือคำถามที่จำเป็นต้องมีการศึกษามากขึ้น หรือก่อให้เกิดสมมติฐานที่จำเป็นต้องมีการทดสอบต่อไป (2)

พารามิเตอร์ที่เกี่ยวข้องใน meta-analysis

ในการรวมผลการวิจัยเข้าด้วยกัน ก่อนอื่นจะต้องคำนวณพารามิเตอร์ที่เหมาะสมสำหรับแสดงผลที่เกิดขึ้นในงานวิจัยแต่ละเรื่องก่อน ซึ่งจะใช้พารามิเตอร์ใดขึ้นอยู่กับลักษณะข้อมูลหรือผลที่วัด หลังจากนั้นจึงเป็นการสังเคราะห์รวมผลที่ได้จากงานวิจัยเหล่านั้นเข้าด้วยกัน ในที่นี้จะขอกล่าวถึงพารามิเตอร์ทางสถิติที่จำเป็นพอสังเขป เพื่อเป็นพื้นฐานสำหรับความเข้าใจหรือการประยุกต์ใช้ต่อไป

ในการเปรียบเทียบประสิทธิผลของยา 2 ชนิด หรือเปรียบเทียบยากับยาหลอก จะวัดผลสุทธิที่เกิดจากการให้ยาที่สนใจศึกษา (เรียกว่า treatment effect หรือ estimate of effect) ได้โดยการเปรียบเทียบผลที่เกิดขึ้นในกลุ่มที่ได้รับยาที่สนใจศึกษา (หรือเรียกว่ากลุ่มทดลอง, treatment group) กับผลที่เกิดขึ้นในกลุ่มที่ได้รับยาเปรียบเทียบ (หรือเรียกว่ากลุ่มควบคุม, control group) เพื่อดูว่าผลที่เกิดขึ้นจากการให้ยาที่เราสนใจศึกษาจะดีหรือแย่กว่ามากน้อยเพียงใดเทียบกับยาเปรียบเทียบ กรณีที่ผลที่สนใจ (outcome) เป็นข้อมูลชนิดต่อเนื่อง (continuous data) เช่น การเปลี่ยนแปลงของระดับความดันโลหิต หรือ การเปลี่ยนแปลงของระดับไขมันในเลือด เป็นต้น พารามิเตอร์ที่ใช้แสดงผลการเปรียบเทียบระหว่างสองกลุ่ม ได้แก่ mean difference และ standardized mean difference (หรือ effect size) (9, 10) กรณีที่ผลที่เกิดขึ้นเป็นไปไม่ได้สองทาง (binary outcomes) เช่น การตาย-การมีชีวิตรอด การเกิดอาการข้างเคียงจากการใช้ยา-ไม่เกิดอาการข้างเคียงจากการใช้ยา พารามิเตอร์ที่ใช้ คือ relative risk (RR) (หรือ risk ratio หรือ rate ratio) absolute risk reduction (ARR) (หรือ risk difference หรือ rate difference (RD)) number needed to treat (NNT) และ odds ratio (OR) (10)

Continuous outcomes

Mean difference

ในกรณีที่ข้อมูลที่ได้จากงานวิจัยมีการกระจายแบบปกติ และงานวิจัยเหล่านั้นใช้เครื่องมือหรือมาตรวัดเดียวกันในการวัดผลที่เกิดขึ้น เช่น การวัดความดันโลหิตโดยมีหน่วยเป็น มิลลิเมตรปรอท จะสามารถแสดงผลที่ได้จากการเปรียบเทียบระหว่างกลุ่มทดลองกับกลุ่มควบคุมได้โดยใช้ค่า mean difference (d_i) ซึ่งเป็นผลต่างระหว่างค่าเฉลี่ยของทั้งสองกลุ่ม มีค่าเท่ากับ (11)

$$d_i = \bar{X}_{1i} - \bar{X}_{0i}$$

โดย $\bar{X}_{1i}, \bar{X}_{0i}$ เป็นค่าเฉลี่ยในงานวิจัยที่ i สำหรับกลุ่มทดลอง และกลุ่มควบคุม ตามลำดับ

Standardised mean difference หรือ effect size

ในกรณีที่ข้อมูลที่ได้จากงานวิจัยมีการกระจายแบบปกติ แต่งานวิจัยแต่ละเรื่องใช้มาตรวัดแตกต่างกันในการวัดผลอย่าง เดียวกัน เช่น ระดับความรุนแรงของโรคทางจิตเวช mania สามารถวัดได้โดยใช้ Manic State Rating Scale, Young Mania Rating Scale, Petterson Scale, Bech-Rafaelsen Mania Scale, Manchester Nurse Rating Scale for Mania และ Clinician-administered Rating Scale for Mania (12) การที่จะนำผลที่ได้มาเปรียบเทียบและรวมเข้าด้วยกัน จำเป็นจะต้องแปลงให้อยู่ในรูปของ standardized mean difference ซึ่งจะให้ค่าที่ไม่มีหน่วยของการวัดเข้ามา เกี่ยวข้อง ค่า effect size (ES_i) นี้สามารถคำนวณได้จากสูตร (9)

$$ES_i = \frac{\bar{X}_{1i} - \bar{X}_{0i}}{S_i}$$

โดย $\bar{X}_{1i}, \bar{X}_{0i}$ เป็นค่าเฉลี่ยของกลุ่มทดลองและกลุ่มควบคุม ในงานวิจัยที่ i และ S_i เป็นส่วนเบี่ยงเบนมาตรฐานรวม (pooled standard deviation) ซึ่งได้จากการคำนวณจากส่วน เบี่ยงเบนมาตรฐานกลุ่มทดลอง (S_{1i}) และกลุ่มควบคุม (S_{0i})

$$S_i = \sqrt{\frac{(n_{1i} - 1)(S_{1i})^2 + (n_{0i} - 1)(S_{0i})^2}{(n_{1i} + n_{0i} - 2)}}$$

โดย n_{1i} และ n_{0i} เป็นจำนวนผู้ป่วยในกลุ่มทดลองและกลุ่ม ควบคุม ตามลำดับ

Binary outcomes

Relative risk, risk ratio หรือ rate ratio (RR)

ในการเปรียบเทียบระหว่างกลุ่มทดลองและกลุ่มควบคุม และผลที่สนใจคือการเกิดเหตุการณ์ (event) ในแต่ละกลุ่ม จะ สามารถแสดงผลที่ได้ในรูปของตาราง 2x2 (13) ดังตัวอย่าง

	เกิดเหตุการณ์	ไม่เกิดเหตุการณ์	รวม
กลุ่มทดลอง	a	c	a+c
กลุ่มควบคุม	b	d	b+d

จากตาราง

$$\begin{aligned} \text{ความน่าจะเป็นของการเกิดเหตุการณ์ในกลุ่มทดลอง } (p_1) \\ = a/a+c \end{aligned}$$

$$\begin{aligned} \text{ความน่าจะเป็นของการเกิดเหตุการณ์ในกลุ่มควบคุม } (p_0) \\ = b/b+d \end{aligned}$$

สัดส่วนของความน่าจะเป็นของการเกิดเหตุการณ์ในสอง กลุ่มที่เปรียบเทียบกัน (p_1/p_0) มีค่าเท่ากับ RR ค่าที่ได้จะบอก ถึงความเสี่ยง (risk) หรือโอกาสที่จะเกิดเหตุการณ์จากการที่ผู้ ป่วยได้รับยาที่สนใจศึกษา โดยจะแสดงในรูปของร้อยละของ ความเสี่ยงจากการได้รับยาเปรียบเทียบ ค่า RR เท่ากับ 1 แสดงว่าไม่มีความแตกต่างกันระหว่างสองกลุ่มที่เปรียบเทียบกัน ถ้าเหตุการณ์ที่เราสนใจเป็นผลในเชิงลบ (undesirable outcome) และค่า RR ที่ได้น้อยกว่า 1 หมายความว่า การเกิด เหตุการณ์นั้นในกลุ่มทดลองน้อยกว่าในกลุ่มควบคุม นั่นคือยาที่ เราสนใจศึกษามีผลลดความเสี่ยงของเหตุการณ์นั้นๆ ตัวอย่าง เช่น RR มีค่าเท่ากับ 0.30 หรือร้อยละ 30 หมายความว่า โอกาสที่จะเกิดเหตุการณ์ในกลุ่มทดลองมีค่าประมาณ 1 ใน 3 ของผู้ที่อยู่ในกลุ่มควบคุม

Absolute risk reduction (ARR)

ARR หรือ RD เป็นผลต่างระหว่างความน่าจะเป็นของการ เกิดเหตุการณ์ในกลุ่มทดลองและกลุ่มควบคุม จากตาราง ARR จะมีค่าเท่ากับ $[b/(b+d)] - [a/(a+c)]$ ค่าที่ได้จะบอกถึง ร้อยละของผู้ป่วยที่เราสามารถป้องกันไม่ให้เกิดเหตุการณ์ที่เป็น ผลในเชิงลบได้โดยการให้ยาที่เราสนใจศึกษา ในกรณีที่การเกิด เหตุการณ์ในกลุ่มทดลองมีมากกว่าในกลุ่มควบคุม ผลต่างที่ได้ อาจเรียกว่า absolute risk increase (ARI) (14) ARR ที่ เท่ากับ 0 แสดงว่าไม่มีความแตกต่างกันระหว่างสองกลุ่ม ตัวอย่าง ARR มีค่าเท่ากับ 0.08 หรือร้อยละ 8 หมายความว่า ถ้า ให้ยาที่เราสนใจศึกษาแก่ผู้ป่วย 100 คน จะป้องกันการเกิดเหตุ การณ์ที่เป็นผลในเชิงลบได้ 8 คน

Number needed to treat (NNT)

จากค่า ARR ที่ได้สามารถนำไปคำนวณหาค่า number needed to treat (NNT) ซึ่งมีค่าเท่ากับส่วนกลับของ ARR ($NNT = 1/ARR$) NNT ถือเป็นพารามิเตอร์อีกตัวหนึ่งที่ใช้ในการประเมินประสิทธิผลของยา โดยค่าที่ได้จะบอกถึงจำนวนผู้ป่วยที่จะต้องได้รับยาที่เราสนใจศึกษา เพื่อที่จะป้องกันการเกิดเหตุการณ์ที่เป็นผลในเชิงลบ (หรือเพื่อให้เกิดผลในเชิงบวก) 1 เหตุการณ์ (15) เมื่อ RR มีค่าไม่เปลี่ยนแปลง NNT จะมีค่าแปรผกผันกับค่าความเสี่ยงในกลุ่มที่ได้รับยาเปรียบเทียบ (เรียกว่า baseline risk) ตัวอย่างเช่น ถ้า baseline risk ลดลงเป็นครึ่งหนึ่ง ค่า NNT จะเพิ่มเป็น 2 เท่า หรือถ้า baseline risk เพิ่มขึ้น 4 เท่า ค่า NNT จะลดลงเป็น 1 ใน 4 ของจำนวนเดิมนั่นคือ ให้ยาที่เราสนใจศึกษากับผู้ป่วยเพียง 1 ใน 4 ของจำนวนผู้ป่วยเดิมก็จะสามารถป้องกันการเกิดเหตุการณ์ในเชิงลบได้ 1 เหตุการณ์ การตัดสินใจทางคลินิกว่าจะเลือกใช้หรือไม่ใช้ยาใดสามารถพิจารณาได้จากค่า ARR และ NNT (16) NNT ที่มีค่าน้อยๆ หมายความว่าเพียงแคเราให้ยาที่สนใจศึกษากับผู้ป่วยเพียงไม่กี่คน ก็จะสามารถป้องกันการเกิดเหตุการณ์ที่ไม่พึงประสงค์ในผู้ป่วยได้ 1 คน ในทางตรงกันข้ามกับ NNT ที่มีค่ามากๆ นั่นคือ ต้องให้ยาที่สนใจศึกษากับผู้ป่วยเป็นจำนวนมากจึงจะป้องกันการเกิดเหตุการณ์เดียวกันนี้ได้ในผู้ป่วย 1 คน

Odds ratio (OR)

OR เป็นสัดส่วนของ odds ระหว่างสองกลุ่มที่เปรียบเทียบกัน สำหรับค่า odds ในแต่ละกลุ่มจะมีค่าเท่ากับความน่าจะเป็นที่จะเกิดเหตุการณ์ ทารด้วยความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ จากตารางข้างต้น OR จะมีค่าเท่ากับ ad/bc ค่า OR ที่เท่ากับ 1 หมายความว่า โอกาสที่จะเกิดเหตุการณ์ที่สนใจในทั้งสองกลุ่มไม่แตกต่างกัน นั่นคือไม่มีความแตกต่างระหว่างสองกลุ่มที่เปรียบเทียบกัน OR จะมีค่าใกล้เคียงกับ RR ในกรณีที่เหตุการณ์ที่สนใจเกิดขึ้นน้อยมาก เช่น อาการอันไม่พึงประสงค์จากการใช้ยาที่พบได้น้อย หรือจะใช้ OR แทน RR ได้ในกรณีที่ไม่สามารถคำนวณ RR ได้ เช่น ใน case-control studies (17)

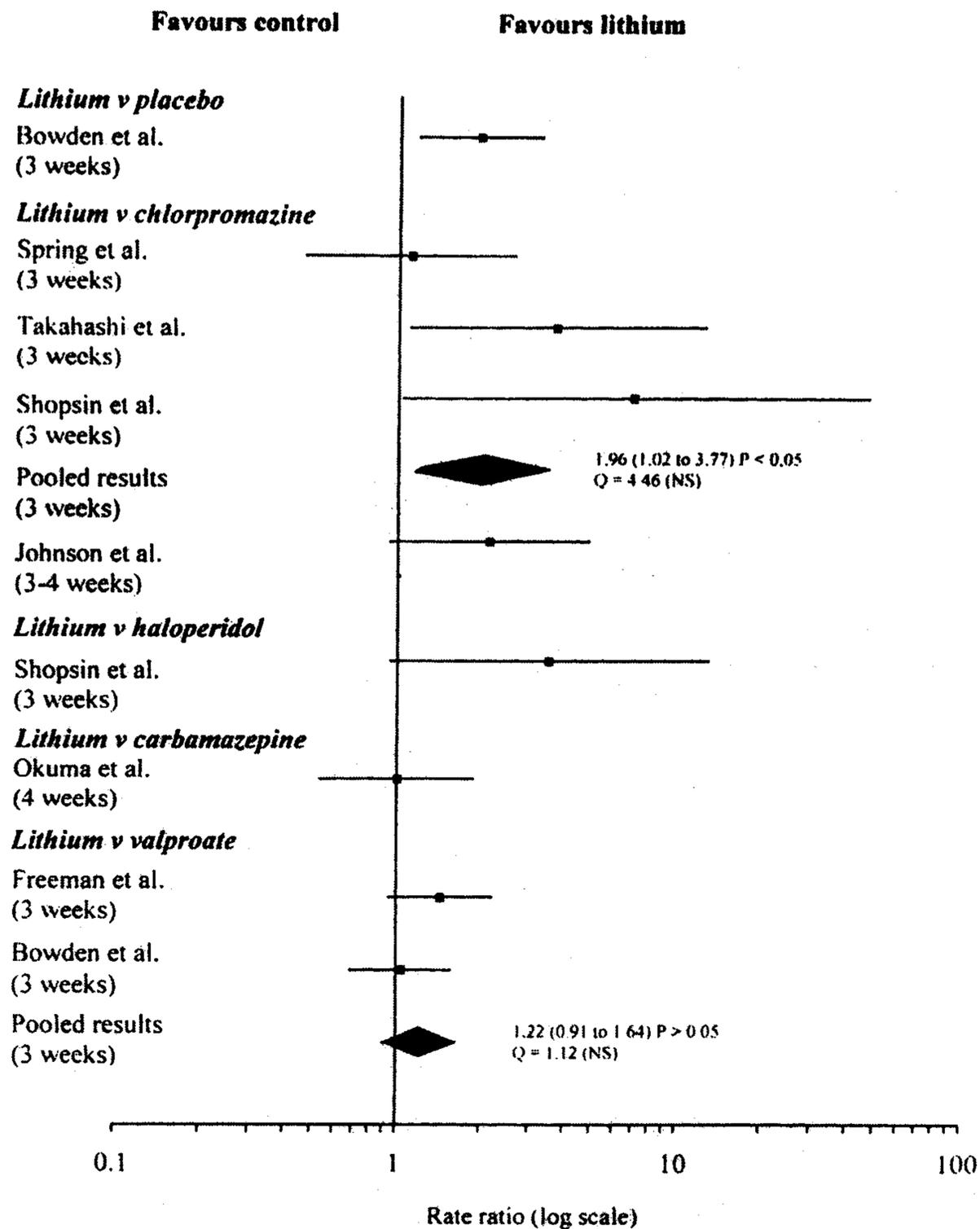
การสังเคราะห์ข้อมูล (Quantitative synthesis)

จากค่าพารามิเตอร์สำหรับแสดงผลที่ได้จากการเปรียบเทียบ

ระหว่างกลุ่มทดลองกับกลุ่มควบคุมตั้งที่กล่าวมาข้างต้น ขึ้นต่อไปคือ การรวมหรือสังเคราะห์ผลที่ได้จากงานวิจัยแต่ละเรื่องเข้าด้วยกัน

การรวมผลการวิจัยเข้าด้วยกันใน meta-analysis สามารถทำได้โดยใช้แบบจำลองทางสถิติ 2 ชนิด คือ fixed effects model และ random effects model โดยใน fixed effects model มีข้อกำหนดว่างานวิจัยที่นำมารวมกันมีค่า treatment effect ที่แท้จริงค่าเดียวกัน และความแตกต่างระหว่างผลการวิจัยมีสาเหตุมาจากความคลาดเคลื่อนแบบสุ่ม (random error) ภายในงานวิจัย ซึ่งถือเป็นความแปรปรวนภายในงานวิจัยเอง (within-study variation) และจะไม่ถูกนำมาพิจารณาในขั้นตอนของการสังเคราะห์ข้อมูล (18) สำหรับ random effects model จะถือว่างานวิจัยที่ถูกนำมารวมกันเป็นตัวอย่างที่ได้จากการสุ่มมาจากประชากรที่แตกต่างกัน treatment effect ที่ได้มีค่าแตกต่างกันไป โดยความแตกต่างที่เกิดขึ้นมีสาเหตุมาจากความแปรปรวนภายในงานวิจัยเอง และความแปรปรวนระหว่างงานวิจัย (between-study variation) ซึ่งมีสาเหตุมาจากปัจจัยอื่นๆ ที่อาจมีผลต่อ treatment effect (19)

การคำนวณผลโดยรวมที่ได้ (pooled estimate) ทำได้โดยการแบ่งน้ำหนักให้กับงานวิจัยแต่ละเรื่องตามความแม่นยำของผลการวิจัยนั้นๆ ซึ่งความแม่นยำนี้จะเป็นส่วนกลับกับค่าความแปรปรวน (variance) ของผลการวิจัย ดังนั้นงานวิจัยที่ผลที่ได้มีความแม่นยำสูง หรืออีกนัยหนึ่งคือมีช่วงความเชื่อมั่น (confidence interval) อยู่ในช่วงแคบๆ ก็จะได้รับน้ำหนักมากกว่างานวิจัยที่มีความแม่นยำต่ำ (นั่นคือมีความแปรปรวนสูง) และเนื่องจากค่าความแปรปรวนของผลการวิจัยแปรผกผันกับจำนวนตัวอย่างที่ศึกษา ดังนั้นในงานวิจัยที่ทำการศึกษาในผู้ป่วยจำนวนมาก ผลที่ได้ก็จะมีค่าความแปรปรวนน้อยและมีความแม่นยำสูง น้ำหนักที่ให้กับงานวิจัยนี้ก็มากกว่าน้ำหนักที่ให้กับงานวิจัยที่ศึกษาในผู้ป่วยจำนวนน้อย (10, 20) การรายงานผลใน meta-analysis มักจะอยู่ในรูปของค่ากลาง (point estimate) และช่วงความเชื่อมั่น 95% การแสดงผลในรูปของกราฟ ที่เรียกว่า forest plot, cochrane plot หรือ blobbogram (รูปที่ 1) จะช่วยให้เห็นภาพชัดเจนและเข้าใจง่าย



รูปที่ 1 Forest plot แสดงค่า RR (95% confidence interval) ของการตอบสนองต่อการรักษาในการเปรียบเทียบประสิทธิผลของ lithium กับ controls (ในที่นี้คือ placebo chlorpromazine haloperidol carbamazepine และ valproate) ในการรักษาผู้ป่วยจิตเวช acute mania (21) เส้นในแนวนอนแสดงผลที่ได้จากงานวิจัยแต่ละเรื่อง จุดที่แบ่งเส้นในแนวนอนคือ point estimate เส้นที่ลากผ่านจุดนี้แสดงถึงช่วงความเชื่อมั่น 95% ซึ่งถ้าเส้นนี้ตัดกับเส้นที่ RR มีค่าเท่ากับ 1 (เส้นในแนวตั้ง ซึ่งเรียกว่า "line of no effect") หมายความว่า treatment effect ที่ได้ไม่มีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 นั่นคือไม่มีความแตกต่างระหว่างยาที่เปรียบเทียบกัน สีเหลี่ยมข้าวหลามตัดแสดงผลรวมที่ได้ (pooled results) จากตัวอย่างจากการเปรียบเทียบระหว่าง lithium กับ chlorpromazine จะเห็นว่า lithium มีประสิทธิผลในการรักษาอาการของ acute mania ดีกว่า chlorpromazine โดยมีค่า RR (95%CI) ของการตอบสนองต่อการรักษาเท่ากับ 1.96 (1.02-3.77) ($p < 0.05$) ค่าสถิติ Q เท่ากับ 4.46 ที่ degree of freedom เท่ากับ 2 ซึ่งได้ค่า p-value > 0.1 (non-significant, NS)

ขึ้นโดยทั่วไป fixed effects model และ random effects model จะให้ค่ากลางที่ใกล้เคียงกัน แต่ random effects model จะให้ค่าช่วงความเชื่อมั่นที่กว้างกว่า แต่ก็ยังไม่มีข้อสรุปว่าวิธีใดเหมาะสมที่สุด ทั้งสองวิธีมีทั้งข้อดีและข้อด้อยในตัวเอง อย่างไรก็ตาม random effects model อาจจะเหมาะสมกว่าในกรณีที่ผลการวิจัยมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (heterogeneous) และไม่สามารถหาสาเหตุของความแตกต่างนั้นได้ (22) ปัจจุบันมีโปรแกรมคอมพิวเตอร์จำนวนมาก ทั้งที่เป็นโปรแกรมสำหรับ meta-analysis โดยเฉพาะ และที่เป็นโปรแกรมทางสถิติทั่วไปที่มีฟังก์ชันสำหรับ meta-analysis รวมอยู่ด้วย ซึ่งบางโปรแกรมสามารถโอนย้ายข้อมูล (download) ได้จากอินเทอร์เน็ตโดยไม่เสียค่าใช้จ่ายใดๆ (23)

การทดสอบความคล้ายคลึงกันของผลการวิจัย (Test of homogeneity)

โดยทั่วไป treatment effect ที่ได้จากงานวิจัยที่ศึกษาปัญหาวิจัยเดียวกันจะมีค่าแตกต่างกันไป ถ้าความแตกต่างนี้มีสาเหตุมาจากความคลาดเคลื่อนแบบสุ่มเพียงสาเหตุเดียว จะถือว่าผลการวิจัยเหล่านั้นไม่แตกต่างกัน (homogeneous) และสามารถรวมกันได้โดยใช้ fixed effects model ในทางตรงกันข้าม ถ้าความแตกต่างที่พบมีค่ามากเกินกว่าที่จะมีสาเหตุมาจากความคลาดเคลื่อนแบบสุ่มเพียงอย่างเดียว แสดงว่าผลการวิจัยเหล่านั้นแตกต่างกันจริง

Test of homogeneity หรือบางคนอาจใช้คำว่า test of heterogeneity เป็นการตรวจสอบความเป็นไปได้ที่ความแตกต่างของผลการวิจัยจะมีสาเหตุมาจากความบังเอิญ (หรือความคลาดเคลื่อนแบบสุ่ม) เพียงอย่างเดียว การทดสอบทางสถิติที่นิยมใช้คือ Q-statistic (20) ซึ่งทดสอบสมมติฐานไร้นัยสำคัญ (null hypothesis) $H_0: \theta_1 = \theta_2 = \dots = \theta_k$ เมื่อ θ_i คือ ค่า treatment effect ที่แท้จริงที่ได้จากงานวิจัยที่ i โดยค่าสถิติ Q จะมีการกระจายตัวคล้ายกับการกระจายตัวของค่าสถิติ χ^2 ที่ degree of freedom เท่ากับ $k-1$ เมื่อ k คือจำนวนของงานวิจัยที่นำมาศึกษา (24) เนื่องจาก Q-statistic มีความไว (sensitivity) ต่ำในการตรวจสอบความแตกต่างของผลการวิจัย นั่นคือ ถึงแม้ว่าค่า Q-statistic ที่ได้จะไม่มีความสำคัญทางสถิติ

แต่ก็อาจเป็นไปได้ที่ผลการวิจัยเหล่านั้นแตกต่างกันจริง ใน meta-analysis จึงแนะนำให้กำหนดระดับนัยสำคัญที่ 0.10 (25) ค่า $p > 0.10$ แสดงว่าผลการวิจัยที่ได้ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ทั้งนี้ไม่ได้เป็นการพิสูจน์ว่าผลที่ได้คล้ายคลึงกัน ค่า $p < 0.10$ แสดงว่าผลการวิจัยที่ได้แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ในกรณีเช่นนี้ควรมีการค้นหาสาเหตุของความแตกต่างต่อไป แต่ถ้าไม่สามารถหาเหตุผลมาอธิบายได้ ควรจะรวมผลการวิจัยเหล่านั้นโดยใช้ random effects model (22)

สาเหตุของความแตกต่างของผลการวิจัย (Causes of heterogeneity)

ความแตกต่างของผลการวิจัยอาจเกิดได้จากความบังเอิญ (เช่น type I error สำหรับ Q-statistic) ความแตกต่างกันในคุณลักษณะและขั้นตอนการดำเนินการวิจัย หรืออาจเกิดจากตัวแปรอื่นๆที่เกี่ยวกับตัวผู้ป่วยเอง รวมทั้งอาจเป็นไปได้ที่ไม่สามารถหาสาเหตุที่แท้จริงของความแตกต่างนั้นได้ (26) สาเหตุที่เกี่ยวข้องกับรูปแบบและวิธีดำเนินการวิจัยที่อาจส่งผลให้ผลการวิจัยแตกต่างกัน ได้แก่ ความแตกต่างกันในเกณฑ์การคัดเลือกผู้ป่วยเข้าร่วมในงานวิจัย หรือถึงแม้จะใช้เกณฑ์เดียวกันในการคัดเลือกผู้ป่วย แต่มีความแตกต่างกันในระดับความรุนแรงของโรคเมื่อเริ่มทำการวิจัย ความแตกต่างกันในแบบแผนการให้ยา (ขนาดยา เวลาที่ให้ยา ระยะเวลาที่ใช้ยา ชื่อการค้า การใช้ยาอื่นร่วมกัน) และสถานที่ทำการวิจัย (หน่วยคนไข้ใน/คนไข้นอก โรงพยาบาลของมหาวิทยาลัย/โรงพยาบาลขนาดเล็ก) รวมทั้งปัจจัยที่เกี่ยวข้องกับการวัดผล (outcome) เช่น การกำหนดขอบเขต ความหมายและลักษณะการวัดผล ระยะเวลาในการติดตามผู้ป่วย และคุณภาพของการวิจัย ล้วนมีส่วนให้ผลการวิจัยที่ได้แตกต่างกันไป (20, 22)

Sensitivity analysis

ใน meta-analysis บางเรื่อง อาจมีการรายงานผลการวิเคราะห์ที่เป็น sensitivity analysis รวมอยู่ด้วย โดยเป็นการตรวจสอบว่าผลโดยรวมที่ได้มีความมั่นคง (robust) มากน้อยเพียงใดเมื่อมีการเปลี่ยนแปลงข้อกำหนด (assumption) หรือกระบวนการในการรวบรวมและสังเคราะห์ผลการวิจัย โดย

เปลี่ยนแปลงข้อมูลไปในลักษณะต่างๆ สิ่งเคราะห์ข้อมูลที่ได้ และเปรียบเทียบผลที่เกิดขึ้นเหล่านั้น วิธีนี้จะบอกได้ว่าข้อสรุปที่ได้มีความมั่นคงเพียงใดต่อการเปลี่ยนแปลงที่เกิดขึ้น ตัวอย่างของ sensitivity analysis ได้แก่ การเปลี่ยนแปลงเกณฑ์การคัดเลือกงานวิจัยที่จะนำมาศึกษา ซึ่งอาจเป็นชนิดของผู้ป่วย การให้ยา หรือตัววัดผลที่สนใจ การรวมเข้าหรือการคัดออกสำหรับงานวิจัยที่ไม่แน่ใจว่าจะมีลักษณะตรงกับเกณฑ์การคัดเลือกงานวิจัยที่จะนำมาศึกษา การคัดออกสำหรับงานวิจัยที่ไม่ได้รับการตีพิมพ์อย่างเป็นทางการในวารสารทางวิชาการ การคัดออกสำหรับงานวิจัยที่มีคุณภาพการดำเนินการวิจัย (methodological quality) ต่ำ และในกรณีที่ม้งานวิจัยขนาดใหญ่เมื่อเทียบกับงานวิจัยอื่นๆ ใน meta-analysis ซึ่งอาจมีผลต่อผลสรุปโดยรวมที่ได้ นอกจากนี้ อาจทำ sensitivity analysis เพื่อเปรียบเทียบผลที่ได้เมื่อใช้วิธีทางสถิติที่ต่างกัน (เช่น fixed effects และ random effects models) (10, 20, 22)

ปัญหาที่พบใน meta-analysis

ถึงแม้ว่าการรวมผลที่ได้จากงานวิจัยหลายๆ เรื่องเข้าด้วยกัน จะทำให้จำนวนตัวอย่างที่ศึกษาเพิ่มมากขึ้น เป็นการเพิ่มความสามารถทางสถิติในการที่จะแสดงให้เห็นถึงความแตกต่างระหว่างกลุ่มที่เปรียบเทียบกันและเพิ่มความแม่นยำของผลที่ได้ แต่ข้อสรุปที่ได้ อาจจะไม่ถูกต้อง (invalid) หรือชี้ไปในทางที่ผิด (misleading) ถ้า meta-analysis นั้นมีวิธีการวิจัยที่ไม่เหมาะสม ในที่นี้จะขอนำเสนอถึงประเด็นที่มีการกล่าวถึงใน meta-analysis

อคติที่เกิดจากการรวบรวมเฉพาะงานวิจัยที่ได้รับการตีพิมพ์ (Publication bias)

เกิดจากการที่งานวิจัยที่แสดงผลที่มีนัยสำคัญทางสถิติมีโอกาที่จะได้รับการตีพิมพ์เผยแพร่ในวารสารทางวิชาการมากกว่างานวิจัยที่แสดงผลที่ไม่มีนัยสำคัญทางสถิติ นอกจากนี้งานวิจัยเหล่านี้ยังมีโอกาที่จะได้รับการตีพิมพ์มากกว่าหนึ่งครั้ง การที่ meta-analysis รวบรวมเฉพาะข้อมูลที่ได้จากผลการวิจัยที่มีการตีพิมพ์เผยแพร่ โดยไม่ได้มีการสืบค้นข้อมูลอย่างเป็นระบบ ครบถ้วนและสมบูรณ์ และไม่ได้สืบค้นไปถึงแหล่งข้อมูลอื่นๆ ที่ไม่ได้ถูกรวมอยู่ในฐานข้อมูลที่มีใช้กันโดยทั่วไปหรือไม่

การจัดทำดัชนีเอาไว้ อาจทำให้ผลโดยรวมที่ได้มีค่าเกินกว่าความเป็นจริง (overestimate) (27, 28)

การรวมผลการวิจัยที่ไม่ได้ตีพิมพ์เผยแพร่ (Inclusion of unpublished data)

วิธีที่จะหลีกเลี่ยงปัญหานี้เนื่องมาจากอคติที่เกิดจากการรวบรวมเฉพาะงานวิจัยที่ได้รับการตีพิมพ์ คือ การนำเอาข้อมูลที่ไม่ได้มีการตีพิมพ์เผยแพร่เข้ามารวมอยู่ใน meta-analysis ด้วย อย่างไรก็ตาม เราไม่สามารถทราบได้ว่าการกระทำดังกล่าวจะเพิ่มหรือลดอคติใน meta-analysis เนื่องจากเราไม่สามารถมั่นใจได้ว่าจะทำการสืบค้นไปถึงงานวิจัยเหล่านั้นได้มากน้อยเพียงใด และถึงแม้ว่าจะทราบแหล่งของข้อมูลแต่ก็อาจมีปัญหาจากการที่คณะผู้วิจัยไม่ยินยอมเผยแพร่ข้อมูลในส่วนนั้น นอกจากนี้ ยังอาจเป็นไปได้ที่งานวิจัยที่ไม่ได้รับการตีพิมพ์เหล่านั้นมีความบกพร่องในระเบียบวิธีวิจัยหรือขั้นตอนการดำเนินการวิจัย การรวมเอาข้อมูลในส่วนนี้ก็อาจจะเป็นการเพิ่มอคติต่อผลรวมที่ได้ (29, 30)

การรวมกันของงานวิจัย (Combinability of studies)

ในทางทฤษฎี เมื่อกำหนดให้ความแตกต่างระหว่างผลการวิจัยมีสาเหตุมาจากความบังเอิญเพียงอย่างเดียว การนำข้อมูลจากงานวิจัยหลายๆ เรื่องมาวิเคราะห์รวมกันจะช่วยเพิ่มความถูกต้องและความแม่นยำของผลที่ได้ ในความเป็นจริงงานวิจัยแต่ละเรื่องจะมีความแตกต่างกัน ไม่ว่าจะเป็นคุณภาพของการวิจัย หรือรูปแบบและการดำเนินการวิจัย เช่น เกณฑ์การคัดเลือกผู้ป่วย ระยะเวลาการศึกษา ตัววัดผล และความแตกต่างกันในแบบแผนการให้ยา เป็นต้น ประเด็นสำคัญในการรวมกันของข้อมูล คือ จะสามารถนำผลที่ได้จากงานวิจัยที่แตกต่างกันมารวมกันได้หรือไม่ การตรวจสอบว่าผลการวิจัยเหล่านั้นแตกต่างกันจริงหรือไม่จึงเป็นสิ่งจำเป็น การนำผลการวิจัยที่แตกต่างกันมารวมกันอาจได้ข้อสรุปที่ไม่ถูกต้อง หรือไม่เหมาะสมและไม่สามารถนำไปใช้ได้ เปรียบเหมือนการนำแอปเปิ้ลมารวมกับส้ม ซึ่งไม่สามารถแปลความหมายของผลที่ได้มา (31) ในทางตรงกันข้ามการนำงานวิจัยที่มีความหลากหลายมารวมกันและถ้ามีการค้นหาสาเหตุของความแตกต่างในผลการวิจัยเหล่านั้นได้ ก็จะช่วยในการแปลผลข้อมูลที่มีอยู่ในขณะนั้นและช่วยการวางแผนการศึกษาวิจัยในขั้นต่อไป

ได้ (27, 32) การค้นหาสาเหตุของความแตกต่างในผลการวิจัยสามารถทำได้โดยการแยกงานวิจัยที่มีลักษณะใกล้เคียงกันออกเป็นกลุ่มย่อยๆ และวิเคราะห์ผลแต่ละกลุ่มแยกกัน (subgroup analysis) หรือใช้วิธีทางสถิติอื่นๆ เช่น meta-regression model หรือ mixed-effect models เป็นต้น (20)

สรุป

ในปัจจุบัน systematic review และ meta-analysis ได้มีบทบาทและความสำคัญเพิ่มมากขึ้นในส่วนที่เกี่ยวข้องกับการดูแลสุขภาพ โดยเฉพาะในการประเมินประสิทธิผลและความปลอดภัยของการใช้ยา การใช้สถิติที่เหมาะสมในการรวมผลการวิจัยเข้าด้วยกัน จะช่วยเพิ่มความสามารถในการแสดงความแตกต่างทางสถิติระหว่างกลุ่มที่เปรียบเทียบกันและเพิ่มความแม่นยำของผลรวมที่ได้ systematic review และ meta-analysis จึงเป็นเครื่องมือสำคัญในการตอบปัญหา และยังมีส่วนช่วยในการวางแผนการวิจัยต่อไปข้างหน้า อย่างไรก็ตาม meta-analysis อาจให้ผลที่ชี้ไปในทางที่ผิดในกรณีที่ meta-analysis นั้นมีแบบแผนการดำเนินการที่ไม่เหมาะสม meta-analysis ที่ต้องการตอบปัญหาในเรื่องเดียวกันอาจให้ข้อสรุปที่แตกต่างกันได้ ดังนั้นจึงจำเป็นที่จะต้องมีการประเมินความน่าเชื่อถือและแปลผลที่ได้ อย่างมีวิจารณญาณก่อนที่จะนำผลที่ได้ไปใช้ต่อไป

เอกสารอ้างอิง

- G. V. Glass. Primary, secondary, and meta-analysis of research. *Educ. Res.* 5: 3-8 (1976).
- K. Dickersin and J. A. Berlin. Meta-analysis: State-of-the-Science. *Epidemiol. Rev.* 14: 154-176 (1992).
- A. Li Wan Po. *Dictionary of Evidence-based medicine*, Radcliffe Medical Press, Oxon, 1998, pp.101-102.
- D. J. Cook, C. D. Mulrow, and R. B. Haynes. Systematic reviews: Synthesis of best evidence for clinical decisions. *Ann. Intern. Med.* 126: 376-380 (1997).
- D. J. Cook, D. L. Sackett, and W. O. Spitzer. Methodologic guidelines for systematic reviews of randomised controlled trials in health care from the Postdam consultation on meta-analysis. *J. Clin. Epidemiol.* 48: 167-171 (1995).
- A. R. Jadad. *Randomised Controlled Trials: A User's Guide*, BMJ Books, London, 2000.
- นงลักษณ์ วิรัชชัย. การวิเคราะห์ที่อภิธาน *Meta-analysis*, จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2542, หน้า 1-12.
- H. S. Sacks, J. Berrier, D. Reitman, V. A. Ancona-Berk, and T. C. Chalmers. Meta-analyses of randomised controlled trials. *N. Engl. J. Med.* 316: 450-455 (1987).
- L. V. Hedges. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Stat.* 6: 107-128 (1981).
- J. Lau, J. P. A. Ioannidis, and C. H. Schmid. Quantitative synthesis in systematic reviews. *Ann. Intern. Med.* 127: 820-826 (1997).
- W. R. Shadish and C. K. Haddock. Combining estimates of effect size. In H. Cooper and L. V. Hedges (eds.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, 1994, pp. 261-281.
- N. Poolsup, A. Li Wan Po, and F. Oyeboode. Measuring mania: A critical appraisal of rating scales. *J. Clin. Pharm. Ther.* 24: 433-443 (1999).
- N. Mantel and W. Haenszel. Statistic aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* 22: 719-748 (1959).
- D. L. Sackett, S. E. Straus, W. S. Richardson, W. Rosenberg, and R. B. Haynes. *Evidence-based Medicine. How to Practice and Teach EMB*, 2nd ed., Churchill Livingstone, China, 2000, pp. 105-153.
- R. J. Cook and D. L. Sackett. The number needed to treat: A clinically useful measure of treatment effect. *BMJ.* 320: 452-454 (1995).
- R. Jaeschke, G. Guyatt, H. Shannon, S. Walter, D. Cook, and N. Heddle. Assessing the effects of treatment: Measuring of association. *Can. Med. Assoc. J.* 152: 351-357 (1995).
- D. G. Altman, J. J. Deeks, and D. L. Sackett. Odds ratios should be avoided when events are common. *BMJ.* 317: 1318 (1998).
- A. Whitehead and J. Whitehead. A general parametric approach to the meta-analysis of

- randomised clinical trials. *Stat. Med.* 10: 1665-1677 (1991).
19. R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Controlled Clin. Trials* 7: 177-188 (1986).
 20. A. J. Sutton, K. R. Abrams, D. R. Jones, T. A. Sheldon, and F. Song. Systematic reviews of trials and other studies. *Health Technol. Assess.* 2: 1-276 (1998).
 21. N. Poolsup, A. Li Wan Po, and I.R. de Oliveira. Systematic overview of lithium treatment in acute mania. *J. Clin. Pharm. Ther.* 25: 139-156 (2000).
 22. J. Deeks, K. S. Khan, F. Song, J. Popay, J. Nixon, and J. Kleijnen. Data synthesis. In K. S. Khan, G. ter Riet, J. Glanville, A. J. Sowden, and J. Kleijnen (eds.), *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying out or Commissioning Reviews, CDR Report Number 4*, 2nd ed., NHS Centre for Review and Dissemination, York Publishing Services, York, 2001, pp. 79-108.
 23. J. A. C. Sterne, M. Egger, and A. J. Sutton. Meta-analysis software. In M. Egger, G.D. Smith, and D. G. Altman, *Systematic Reviews in Health Care. Meta-analysis in Context*, BMJ Books, Cornwall, 2001, pp. 336-346.
 24. W. G. Cochran. The combination of estimates from different experiments. *Biometrics* 10: 101-129 (1954).
 25. J. L. Fleiss. Analysis of data from multiclinic trials. *Controlled Clin. Trials* 7: 267-275 (1986).
 26. K. R. Bailey. Inter-study differences -- how should they influence the interpretation and analysis of results. *Stat. Med.* 6: 351-360 (1987).
 27. J. A. Berlin. The use of meta-analysis in pharmacoepidemiology. In B. L. Strom (ed.), *Pharmacoepidemiology*, 3rd ed., John Wiley & Sons, Midsomer Norton, 2000, pp. 633-659.
 28. C. D. Naylor. Meta-analysis and the meta-epidemiology of clinical research. *BMJ.* 315: 617-619 (1997).
 29. G. D. Smith and M. Egger. Meta-analysis. Unresolved issues and future developments. *BMJ.* 316: 221-225 (1998).
 30. D. J. Cook, G. H. Guyatt, G. Ryan, J. Clifton, L. Buckingham, A. Willan, W. McIlroy, and A. D. Oxman. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA.* 269: 2749-2753 (1993).
 31. H. J. Eysenck. Meta-analysis and its problems. *BMJ.* 309: 789-792 (1994).
 32. A. Li Wan Po. Evidence-based pharmacotherapy. *Pharm. J.* 256: 308-312 (1996).