

การประเมินงานวิจัยนิตการทดลองแบบสุ่มที่มีการควบคุม: อภินิหารเครื่องมือแคลสพี กลุ่มพรางของการใช้แคลสพี และเครื่องมือประเมินชนิดอื่นๆ

ณัฐพัฒน์ ทองน้อย* ณัฐ นาเอก, ภบ (บริหารเภสัชกรรม), MSc, PhD (Clinical Epidemiology)**

บทคัดย่อ

การประเมินความน่าเชื่อถือของวรรณกรรมปฐมภูมิ โดยเฉพาะอย่างยิ่งงานวิจัยรูปแบบสุ่มที่มีการควบคุม เป็นทักษะการคิดเชิงวิพากษ์ที่จำเป็นสำหรับบุคลากรทางการแพทย์เพื่อใช้ประกอบการดูแลผู้ป่วย บทความฉบับก่อนหน้า ได้แนะนำเครื่องมือประเมินความน่าเชื่อถือของงานวิจัยรูปแบบสุ่มที่มีการควบคุมที่มีชื่อว่าแคลสพีซึ่งประกอบด้วย 11 ข้อคำถามเพื่อตอบคำถามที่สำคัญสามข้อ 1) ผลลัพธ์ของการศึกษามีความถูกต้องหรือไม่ 2) ผลลัพธ์ของการศึกษาเป็นอย่างไร และ 3) จะประยุกต์ใช้ผลลัพธ์ของการศึกษากับผู้ป่วยได้อย่างไร บทความฉบับนี้จะเป็นการอธิบายถึงรายละเอียดของเครื่องมือแคลสพีที่มีการปรับปรุงล่าสุด โดยได้เปรียบเทียบกับฉบับก่อนหน้า นอกจากนี้ยังแสดงให้เห็นถึงข้อควรระวังหรือหลุมพรางในการใช้เครื่องมือดังกล่าวจากประสบการณ์ของผู้เขียน ตลอดจนมีการแนะนำเครื่องมืออื่น ๆ สำหรับใช้ประเมินความน่าเชื่อถือของงานวิจัยรูปแบบสุ่มที่มีการควบคุม ทั้งนี้มีวัตถุประสงค์หลักเพื่อแนะนำเทคนิคสำหรับสนับสนุนการตัดสินใจ เพื่อให้ผู้อ่านสามารถนำไปประยุกต์ใช้ในการปฏิบัติงานและพัฒนาการดูแลผู้ป่วยผ่านหลักฐานเชิงประจักษ์ทางการแพทย์ต่อไป

คำสำคัญ

การวิจัยแบบสุ่มที่มีการควบคุม การประเมินความน่าเชื่อถือ เครื่องมือแคลสพี เทคนิคสนับสนุนการตัดสินใจ การวางแผนดูแลผู้ป่วย

*นิสิตชั้นปีที่ 6 คณะเภสัชศาสตร์ มหาวิทยาลัยพะเยา

**สาขาวิชาบริหารเภสัชกรรม คณะเภสัชศาสตร์ มหาวิทยาลัยพะเยา

ติดต่อเกี่ยวกับบทความ : ณัฐ นาเอก E-mail : nat.na@up.ac.th

วันที่รับเรื่อง : 5 ตุลาคม 2564 วันที่ส่งแก้ไข : 14 พฤศจิกายน 2564 วันที่ตีพิมพ์ : 27 ธันวาคม 2564

CRITICAL APPRAISAL OF A RANDOMISED CONTROLLED TRIAL: AN UPDATE CASP TOOL, THE PITFALLS OF USING CASP, AND THE OTHER CRITICAL APPRAISAL TOOLS

Poomipat Thongnoi*, Nat Na-Ek, PharmD (Pharm Care), MSc, PhD (Clinical Epidemiology)**

ABSTRACT

Critical appraisal of primary literature, primarily randomised controlled trials (RCT), is a critical thinking skill essential for healthcare professionals to support patient care. The previous article introduced a critical appraisal tool for RCT called the critical appraisal skill programmed (CASP). It is composed of 11 items that answer three critical questions: 1) Are the results valid? 2) What are the results? and 3) How can I apply the results to my patients? In this article, the details of the updated CASP tool are explained and compared to the previous version. We also illustrate the caveats or pitfalls of using the CASP tool from our experience. Additionally, other critical appraisal tools for RCT will be introduced. Our main objective is to provide techniques for supporting decision-making skills that readers can apply to their work and improve patient care through evidence-based medicine.

KEYWORDS

randomised controlled trial, critical appraisal, critical appraisal skill programme (CASP), decision support techniques, patient care planning

*6th year pharmacy student, School of Pharmaceutical Sciences, University of Phayao

**Division of Pharmaceutical Care, School of Pharmaceutical Sciences, University of Phayao

Corresponding Author : Nat Na-EK E-mail : nat.na@up.ac.th

Accepted date : 5 October 2021 Revise date : 14 November 2021 Publish date : 27 December 2021

บทนำ

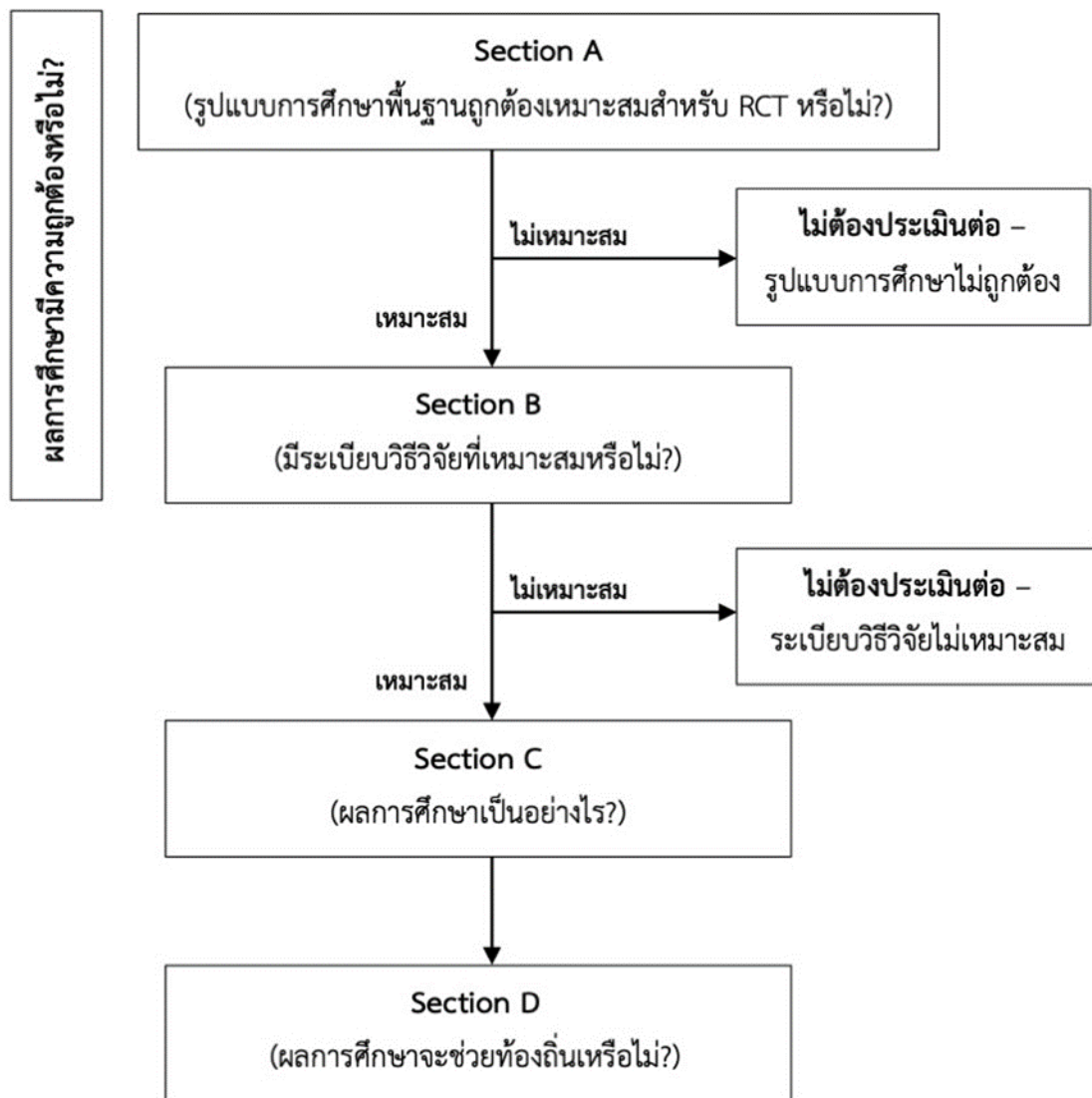
ในบทความก่อนหน้านี้ผู้เขียนได้อธิบายถึงการใช้เครื่องมือแคลสปี (Critical Appraisal Skill Programme; CASP หากออกเสียงตามหลักสากลจะออกเสียงว่า “แคส” แต่ในที่นี้ขอใช้ว่า “แคลสปี” เพื่อให้สอดคล้องกับการสะกดและงานเขียนก่อนหน้านี้) เพื่อใช้ประเมินความน่าเชื่อถือของงานวิจัยรูปแบบสุ่มที่มีการควบคุม (randomised controlled trials; RCT)¹ ซึ่งในปี พ.ศ.2563 ได้มีการปรับปรุงเครื่องมือดังกล่าวโดยการเพิ่มข้อคำถามและปรับรวมคำถามบางข้อเข้าด้วยกัน ส่งผลให้สามารถประเมินคุณภาพของ RCT ได้ครอบคลุมยิ่งขึ้น

บทความฉบับนี้ประกอบด้วยเนื้อหาสามส่วน ส่วนแรกเป็นการอธิบายเครื่องมือแคลสปีฉบับปรับปรุง พ.ศ.2563 โดยได้เปรียบเทียบรายละเอียดการปรับปรุงกับเครื่องมือแคลสปีฉบับก่อนหน้า ส่วนที่สองจะอธิบายถึงหลุมพรางหรือข้อควรระวังในการใช้เครื่องมือแคลสปี ซึ่งรวบรวมขึ้นจากประสบการณ์การสอนและการสังเกตการนำเสนอของงานสังสรรค์วรรณกรรม (journal club) ของนิตเภสัชศาสตร์ ชั้นปีที่ 6 ในระหว่างการศึกษาฝึกปฏิบัติงานวิชาชีพเภสัชกรรม ส่วนสุดท้ายเป็นข้อมูลเกี่ยวกับเครื่องมืออื่นๆ ที่ใช้เพื่อประเมินความน่าเชื่อถือของงานวิจัยรูปแบบสุ่มที่มีการควบคุม ทั้งนี้ เพื่อสนับสนุนแนวคิดว่าการประเมินความน่าเชื่อถือของงานวิจัย ไม่จำเป็นต้องถูกจำกัดกรอบแนวคิดไว้ด้วยเครื่องมือใดเครื่องมือหนึ่งเพียงชนิดเดียว เนื่องจากหัวใจของการประเมินความน่าเชื่อถือของงานวิจัย เป็นการฝึกกระบวนการคิดเชิงวิพากษ์อย่างเป็นระบบ เพื่อให้สามารถดึงข้อมูลจากงานวิจัยไปใช้ดูแลผู้ป่วยได้อย่างเหมาะสมบนพื้นฐานของหลักฐานเชิงประจักษ์ทางการแพทย์

เครื่องมือแคลสปีฉบับปรับปรุง พ.ศ.2563

ดังที่ได้อธิบายไปในบทความฉบับก่อนหน้านี้¹ ว่าการประเมินงานวิจัยประกอบด้วย 4 ขั้นตอนที่สำคัญ ได้แก่ 1) สรุปงานวิจัย 2) ทบทวนเรื่องโรคและข้อมูลยาที่เกี่ยวข้อง 3) ประเมินงานวิจัย และ 4) สรุปผลจากการประเมินงานวิจัยว่า ผลลัพธ์ของการศึกษามีความตรงหรือไม่ (Are the results valid?) ผลลัพธ์ของการศึกษาเป็นอย่างไร (What are the results?) และจะประยุกต์ใช้ผลลัพธ์นี้กับผู้ป่วยได้อย่างไร (How can I apply the results to my patients?) (หรือผลการศึกษาจะช่วยท้องถิ่นหรือไม่; Will the results help locally?)

สำหรับเครื่องมือแคลสปีฉบับปรับปรุง พ.ศ.2563 ยังคงมี 11 ข้อคำถาม โดยจะให้ประเมินว่าผู้วิจัย “ได้ทำ (Yes)” “ไม่ได้ทำ (No)” หรือ “ไม่สามารถระบุได้ (Can't tell)” เช่นเดิม แต่ได้แบ่งคำถามเป็น 4 ส่วน (จากเดิม 3 ส่วน) ได้แก่ A) รูปแบบการศึกษาพื้นฐานมีความถูกต้องสำหรับ RCT หรือไม่ (Is the basic study design valid for an RCT?) B) การศึกษามีระเบียบวิธีวิจัยที่ดีหรือไม่ (Was the study methodologically sound?) C) ผลการศึกษเป็นอย่างไร (What are the results?) และ D) ผลการศึกษาจะช่วยท้องถิ่นนั้น ๆ หรือไม่ (Will the results help locally?) โดยผู้อ่านจะต้องประเมินสองส่วนแรก (A-B) ก่อน หากไม่ผ่านการประเมินอย่างน้อยส่วนใดส่วนหนึ่ง ก็ไม่จำเป็นต้องประเมินในส่วนต่อไป (รูปที่ 1)



รูปที่ 1 องค์ประกอบของเครื่องมือแคลสปีฉบับปรับปรุง พ.ศ.2563

รายละเอียดของแคลสปีฉบับปรับปรุง พ.ศ.2563 เปรียบเทียบกับฉบับก่อนหน้า (ตารางที่ 1) กล่าวโดยสรุป การประเมินความตรงภายในของผลการศึกษาในแคลสปีฉบับปรับปรุงจะคล้ายกับฉบับก่อนหน้า แต่ข้อที่คำถามเรื่องการปกปิด (blinding หรือ masking) จะให้ประเมินแยกจากผู้เข้าร่วมงานวิจัย นักวิจัย และผู้ประเมินผลลัพธ์ถูกปกปิดหรือไม่ นอกจากนี้ส่วนของผลลัพธ์การศึกษาแคลสปีฉบับปรับปรุงได้เพิ่มคำถามเกี่ยวกับความสมบูรณ์ครบถ้วนของการรายงานผลลัพธ์การวิจัย ซึ่งจะเพิ่มการพิจารณาการคำนวณขนาดตัวอย่างและการเลือกใช้

สถิติทดสอบ นอกจากนี้ในส่วนการประยุกต์ใช้ผลลัพธ์การวิจัยกับผู้ป่วย แคลสปีฉบับปรับปรุงยังเพิ่มคำถามเกี่ยวกับความคุ้มค่าของสิ่งแทรกแซง (intervention) ใหม่เปรียบเทียบกับสิ่งแทรกแซงที่ใช้อยู่ในปัจจุบัน ซึ่งเป็นการเพิ่มมุมมองเชิงบริหารและนโยบายเข้ามาร่วมในการประเมินด้วย ส่งผลให้ประเมินได้ครอบคลุมหลากหลายมิติมากขึ้น

ตารางที่ 1 เปรียบเทียบเครื่องมือแคลสปีฉบับเก่าและฉบับปรับปรุง พ.ศ. 2563²

แคลสปีฉบับเก่า (พ.ศ.2561)	แคลสปีฉบับปรับปรุง พ.ศ.2563	หมายเหตุ
Section A: Are the results of the trial valid? 1. Did the trial address a clearly focused issue? 2. Was the assignment of patients to treatments randomised? 3. Were all of the patients who entered the trial properly accounted for at its conclusion?	Section A: Is the basic study design valid for a randomised controlled trial? 1. Did the study address a clearly focused research question? 2. Was the assignment of participants to interventions randomised? 3. Were all of participants who entered the study accounted for at its conclusion?	ฉบับปรับปรุงได้แยกการประเมินความตรงภายในออกเป็น 2 ส่วน (A และ B) คงเดิม คงเดิม คงเดิม
4. Were patients, health workers and study personnel 'blind' to treatment? 5. Were the groups similar at the start of the trial? 6. Aside from the experimental intervention, were the groups treated equally?	Section B: Was the study methodologically sound? 4. Were the participants, the investigators, the people assessing/analysing outcomes 'blinded'? 5. Were the study groups similar at the start of the randomised controlled trial? 6. Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)?	แยกเป็นส่วนใหม่ในฉบับปรับปรุง คล้ายเดิม แต่ฉบับปรับปรุงจะพิจารณาแยกคนแต่ละกลุ่มในงานวิจัย คงเดิม คงเดิม

ตารางที่ 1 เปรียบเทียบเครื่องมือแคสพิฉบับเก่าและฉบับปรับปรุง พ.ศ. 2563²

แคสพิฉบับเก่า (พ.ศ.2561)	แคสพิฉบับปรับปรุง พ.ศ.2563	หมายเหตุ
Section B: What are the results? 7. How large was the treatment effect? 8. How precise was the estimates of the treatment effect?	Section C: What are the results? 7. Were the effects of intervention reported comprehensively? 8. Was the precision of the estimate of the intervention or treatment effect reported? 9. Do the benefits of the experimental intervention outweigh the harms and costs?	คงเดิม (แต่เปลี่ยนเป็นส่วน C) เป็นคำถามใหม่ที่ถูกเพิ่มเข้ามา โดยข้อ 7. เดิมได้ถูกนำไปรวมกับข้อ 9. คล้ายเดิม แต่เปลี่ยนจากคำถาม how เป็น was เพื่อให้ตอบได้ง่ายขึ้น ย้ายมาจากส่วน C ในฉบับก่อนหน้า และเพิ่มการพิจารณาขนาดผลลัพธ์ (effect size) ในข้อนี้
Section C: Will the results help locally? 9. Can the results be applied to the local population, or in your context? 10. Were all clinically important outcomes considered? 11. Are the benefits worth the harms and costs?	Section D: Will the results help locally? 10. Can the results be applied to your local population/ in your context? 11. Would the experimental intervention provide greater value to the people in your care than any of the existing intervention?	คงเดิม (แต่เปลี่ยนเป็นส่วน D) คล้ายเดิมแต่เพิ่มการพิจารณาความครอบคลุมของการรายงานผลลัพธ์ที่สำคัญทางคลินิก เป็นคำถามที่ถูกเพิ่มเข้ามาใหม่ ถูกย้ายไปส่วน C ในฉบับปรับปรุง

ผู้อ่านสามารถเข้าถึงแบบประเมินแคลสป์ฉบับปรับปรุง พ.ศ.2563 ได้จากเว็บไซต์ <https://casp-uk.net/casp-tools-checklists/> 2 โดยคำถามแต่ละข้อมีรายละเอียดดังต่อไปนี้

ส่วนที่หนึ่ง: Are the results valid?
(ผลการศึกษามีความตรงหรือไม่)

รายละเอียดการประเมินความตรงของงานวิจัยรูปแบบ RCT ของเครื่องมือแคลสป์ฉบับปรับปรุง พ.ศ.2563 คล้ายคลึงกับฉบับก่อนหน้า จึงไม่ขออธิบายรายละเอียดในบทความฉบับนี้ ผู้ที่สนใจขอแนะนำให้ศึกษาเพิ่มเติมจากเอกสารอ้างอิง¹

Section A: Is the basic study design valid for an RCT? (รูปแบบการศึกษาพื้นฐานมีความถูกต้องเหมาะสมสำหรับ RCT หรือไม่)

- 1) Did the study address a clearly focused research question? (การศึกษาได้ระบุถึงคำถามงานวิจัยไว้อย่างชัดเจนหรือไม่)
- 2) Was the assignment of participants to interventions randomised? (มีการสุ่มให้สิ่งแทรกแซงแก่ผู้เข้าร่วมการศึกษาหรือไม่)
- 3) Were all of participants who entered the study accounted for at its conclusion? (ข้อสรุปของการศึกษาได้นับรวมผู้เข้าร่วมการศึกษาทุกรายหรือไม่)

Section B: Was the study methodologically sound? (มีระเบียบวิธีวิจัยที่เหมาะสมหรือไม่)

- 4) Were the participants, the investigators, the people assessing/analysing outcomes' blinded? (มีการปกปิดผู้เข้าร่วมการศึกษา แพทย์ผู้ดูแลผู้ป่วย ตลอดจนผู้ประเมิน/วิเคราะห์ผลลัพธ์ของการศึกษาหรือไม่)

- 5) Were the study groups similar at the start of the randomised controlled trial? (แต่ละกลุ่มที่ศึกษามีลักษณะที่คล้ายคลึงกันตั้งแต่เริ่มต้นหรือไม่)
- 6) Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)? (นอกเหนือจากสิ่งแทรกแซงในการทดลองแล้ว ผู้ป่วยแต่ละกลุ่มได้รับการรักษาในระดับเดียวกันหรือไม่ นั่นคือผู้เข้าร่วมการศึกษาถูกรักษาเท่าเทียมกันหรือไม่)

ส่วนที่สอง: What are the results?
(ผลการศึกษาเป็นอย่างไร)

Section C: What are the results? (ผลการศึกษาเป็นอย่างไร)

- 7) Were the effects of intervention reported comprehensively? (ผลของสิ่งแทรกแซงได้ถูกรายงานไว้อย่างครบถ้วนหรือไม่)

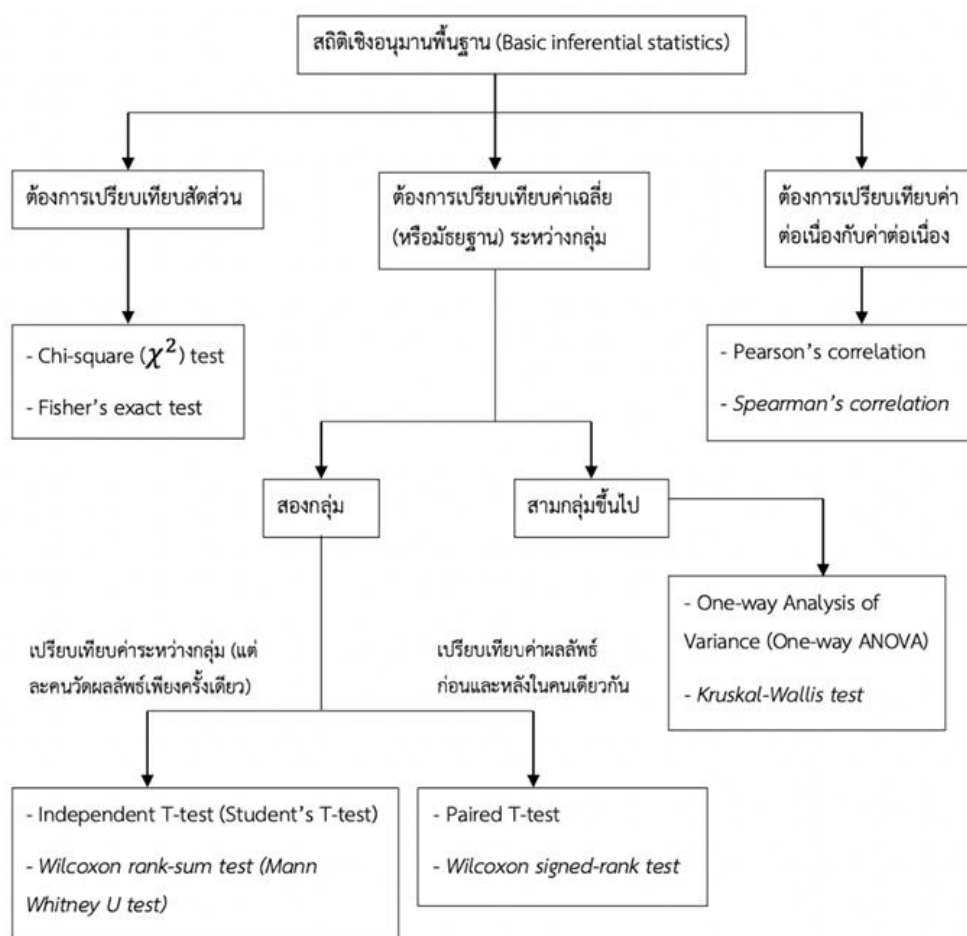
ข้อนี้ให้พิจารณาความสมบูรณ์ของการรายงานผลลัพธ์ในการศึกษา ซึ่งพิจารณาจาก 1) มีการคำนวณขนาดตัวอย่าง 2) มีการระบุชนิดของผลลัพธ์ที่วัดไว้อย่างชัดเจน 3) กรณีที่ผลลัพธ์เป็นข้อมูลแบบแบ่งสอง (binary outcomes) ผู้วิจัยควรนำเสนอทั้งค่าสัมพัทธ์ เช่น relative risk reduction และค่าสัมบูรณ์ เช่น absolute risk reduction 4) ผู้วิจัยควรรายงานค่าของผลลัพธ์ในทุกช่วงเวลาติดตาม 5) ระบุถึงรายละเอียดของข้อมูลสูญหายโดยให้พิจารณาต่อว่าหากมีข้อมูลสูญหายจะส่งผลกระทบต่อผลลัพธ์หรือไม่ 6) วิธีการวิเคราะห์ทางสถิติ และ 7) มีการรายงานค่า p-values ร่วมด้วย

จะเห็นว่าในข้อนี้มีการพิจารณาที่ละเอียด แต่จากประสบการณ์ของผู้เขียนพบว่างานวิจัยส่วนใหญ่มักไม่ได้รายงานข้อมูลเหล่านี้ไว้อย่างครบถ้วน เช่น ผลลัพธ์ในแต่ละช่วงเวลาติดตาม มักไม่ค่อยมีการรายงาน เป็นต้น ผู้เขียนจึงขอแนะนำให้พิจารณาในสามประเด็นที่สำคัญเป็นอย่างน้อย ได้แก่ 1) มีการคำนวณขนาดตัวอย่าง 2) เลือกใช้สถิติทดสอบได้เหมาะสมกับชนิดของผลลัพธ์ที่ต้องการวัด (รูปที่ 2)

การวิเคราะห์ถดถอย (regression analysis) เป็นสถิติที่พบบ่อยในงานวิจัยรูปแบบสุ่มที่มีการควบคุม ซึ่งหลักการเลือกชนิดการถดถอยนั้นจะขึ้นกับชนิดของตัวแปรผลลัพธ์ (outcome-, dependent-,

หรือ Y-variable) ดังแสดงในตารางที่ 2 โดยตัวแปรต่อเนื่อง (continuous variable) คือตัวแปรที่ได้จากการวัด เช่น ค่าผลตรวจทางห้องปฏิบัติการ ค่าน้ำหนัก ค่าส่วนสูง เป็นต้น

ตัวแปรแบ่งกลุ่มสอง (binary variable) คือตัวแปรที่มีสองกลุ่ม เช่น การเกิดโรค (เกิด/ไม่เกิด) การตาย (ตาย/ไม่ตาย) การหายจากโรค (หาย/ไม่หาย) เป็นต้น ส่วนตัวแปรนับ (count variable) คือตัวแปรที่ได้จากการนับจำนวน เช่น จำนวนครั้งของการเกิดภาวะหอบกำเริบในช่วง 1 ปี (ครั้ง) จำนวนฟันผุ (ซี่) จำนวนครั้งของการเกิดภาวะภูมิไวเกินจากการได้รับยาเคมีบำบัดตลอดระยะเวลา 12 cycles (ครั้ง) หรือจำนวนวันนอนโรงพยาบาล (วัน) เป็นต้น



รูปที่ 2 การเลือกใช้สถิติเชิงอนุมานพื้นฐาน (basic inferential statistics)³

หมายเหตุ สัญลักษณ์ตัวเอียง คือ non-parametric tests

ตารางที่ 2 การเลือกใช้การวิเคราะห์ถดถอย (regression analysis) ที่พบบ่อย³

ชนิดของตัวแปรผลลัพธ์	ชนิดการถดถอย	ขนาดผลลัพธ์ที่ได้
ตัวแปรต่อเนื่อง (continuous)	Linear regression	Mean difference
ตัวแปรแบ่งกลุ่มสอง (binary)	Logistic regression	Odds ratio (OR)
	Binomial regression	Risk ratio (Relative risk; RR)
ตัวแปรนับ (count)	Poisson regression	Incidence rate ratio (IRR)
ตัวแปรแบ่งกลุ่มสองที่นำระยะเวลา	Cox-proportional hazards	Hazard ratio (HR)
ติดตามมาพิจารณา (time-to-event)	regression	

8) Was the precision of the estimate of the intervention or treatment effect reported? (มีการรายงานความแม่นยำของขนาดผลลัพธ์ของสิ่งแทรกแซงหรือไม่)

ข้อนี้ให้พิจารณาว่าผู้วิจัยได้รายงานช่วงความเชื่อมั่น (confidence interval) ของผลลัพธ์จากการศึกษาร่วมด้วยหรือไม่ หากมีการรายงานก็สามารถประเมินว่า “Yes” ได้ทันที

9) Do the benefits of the experimental intervention outweigh the harms and costs? (ประโยชน์ที่ได้รับจากสิ่งแทรกแซงที่ศึกษาคู่ค่ากับอันตรายและต้นทุนที่อาจเพิ่มขึ้นหรือไม่)

ในข้อนี้มีสองประเด็นต้องพิจารณา คือ 1) ขนาดของผลลัพธ์และ 2) การเปรียบเทียบค่า number needed to treat (NNT) และค่า number needed to harm (NNH) ซึ่งวิธีการคำนวณ NNT และ NNH ได้ถูกอธิบายไว้อย่างละเอียดในบทความก่อนหน้านี้¹

ในแง่ของขนาดของผลลัพธ์มีสิ่งที่พึงระลึกไว้เสมอ คือ “นัยสำคัญทางสถิติ (statistically significant) ไม่เท่ากับนัยสำคัญทางคลินิก (clinically significant) และนัยสำคัญทางคลินิกไม่สามารถบอกได้จากค่า p-value” เนื่องจากการเพิ่มขนาดตัวอย่างให้ใหญ่ขึ้นอาจทำให้พบความแตกต่างแม้เพียงเล็กน้อย

แม้จะไม่มีนัยสำคัญทางคลินิกก็ตามเรียกว่าเป็นหลุมพรางจากกลุ่มตัวอย่างขนาดใหญ่ เช่น การศึกษาผลของยาลดความดันโลหิตชนิดใหม่เทียบกับยามาตรฐาน หากทดสอบในคนจำนวน 10,000 คน พบว่ายาใหม่มีประสิทธิภาพในการลดความดันโลหิตค่าบน (systolic blood pressure) ลง 0.5 mmHg (95% CI 0.1 to 0.9, $P < 0.001$) ในทางปฏิบัติย่อมไม่มีแพทย์ท่านใดสั่งใช้ยาดังกล่าว เพราะความดันโลหิตค่าบนที่ลดลงเพียง 0.5 mmHg ไม่มีนัยสำคัญทางคลินิก เป็นต้น กล่าวได้อีกนัยหนึ่งว่านัยสำคัญทางสถิติสามารถเปลี่ยนแปลงได้ ขึ้นกับขนาดของกลุ่มตัวอย่างที่ทำการศึกษา โดยอาจไม่สัมพันธ์กับนัยสำคัญทางคลินิก แต่ทั้งนี้ผลลัพธ์ที่จะมีนัยสำคัญทางคลินิกได้จะต้องแสดงถึงนัยสำคัญทางสถิติก่อนเสมอ

ผลการศึกษาที่ไม่พบนัยสำคัญทางสถิติอาจเป็นผลที่มาจาก negative study หรือเป็นผลจากอำนาจทางสถิติไม่เพียงพอที่จะตรวจพบความแตกต่างที่มี (underpower of statistics) ซึ่งการศึกษาที่เป็น negative study จะบ่งชี้ว่าไม่จำเป็นต้องศึกษาเพิ่มเติมเนื่องจากสิ่งแทรกแซงที่ศึกษานั้นไม่มีประโยชน์ทางคลินิกในขณะที่การศึกษาที่มีอำนาจทางสถิติต่ำ (underpowered study) ควรศึกษาเพิ่มเติมในกลุ่มตัวอย่างขนาดใหญ่ขึ้น

(หรือติดตามยาวนานขึ้น) เนื่องจากสิ่งแทรกแซงที่ศึกษาอาจมีประโยชน์ทางคลินิกได้

การพิจารณาว่าการศึกษาใดเป็น negative หรือ underpowered study นั้น พิจารณาจากช่วงความเชื่อมั่นและค่านัยสำคัญทางคลินิก (minimum clinical significance) ตัวอย่างเช่น หากผู้วิจัยกำหนดว่าค่า $RR > 2.00$ มีนัยสำคัญทางคลินิกแล้ว (มักสังเกตได้จากค่าที่ใช้คำนวณขนาดตัวอย่าง) ผลลัพธ์ของการศึกษามี 95%CI อยู่ในช่วง 0.50 ถึง 1.50 แสดงว่าการศึกษาดังกล่าวเป็น negative study เพราะช่วงความเชื่อมั่นไม่ครอบคลุมค่านัยสำคัญทางคลินิกต่อให้ผู้วิจัยเพิ่มขนาดตัวอย่างจนช่วงความเชื่อมั่นแคบลงและไม่คร่อมค่า 1.00 (มีนัยสำคัญทางสถิติ) ก็จะไม่ยังมีนัยสำคัญทางคลินิกอยู่ดี แต่ถ้าผลลัพธ์ของการศึกษามี 95%CI อยู่ในช่วง 0.50 to 2.50 ก็อาจพิจารณาว่าเป็น underpowered study ได้ เนื่องจากช่วงความเชื่อมั่นครอบคลุมค่า 2.00 ที่หากเพิ่มขนาดตัวอย่าง อาจทำให้ช่วงความเชื่อมั่นแคบลงจนมีนัยสำคัญทางสถิติและยังมีนัยสำคัญทางคลินิกไปด้วย เป็นต้น

อนึ่ง ผู้อ่านระวังสับสนระหว่างคำว่า negative study และ negative association คำว่า negative study หมายถึง ผลการศึกษาที่ไม่พบทั้งนัยสำคัญทางสถิติและนัยสำคัญทางคลินิก ในขณะที่ negative association แปลว่าความสัมพันธ์เชิงลบหมายความว่าถ้าสัมผัสปัจจัยเพิ่มขึ้น ความเสี่ยงต่อการเกิดผลลัพธ์จะลดลง (หรือกลับกัน) เช่น การมีระดับ high-density lipoprotein cholesterol (HDL-C) ที่เพิ่มขึ้นจะสัมพันธ์กับความเสี่ยงต่อการเกิดโรคหัวใจ และหลอดเลือดที่ลดลง บางครั้งอาจเรียกความสัมพันธ์เชิงลบว่าเป็นความสัมพันธ์เชิงผกผัน (inverse association)

หากพิจารณาเครื่องมือแคลสป์ฉบับปรับปรุงจะพบว่า คำถามที่ 9 นี้ครอบคลุมการพิจารณาขนาดผลลัพธ์ไว้ด้วย ซึ่งคล้ายกับคำถามข้อ 7 ในแคลสป์ฉบับก่อนหน้า (How large was the treatment effect?) กล่าวโดยสรุป นอกจากจะต้องพิจารณาความคุ้มค่าประสิทธิผลและความปลอดภัยของสิ่งแทรกแซงแล้วยังต้องพิจารณานัยสำคัญทางคลินิกของผลลัพธ์ร่วมด้วยเสมอ

ส่วนที่สาม: How can I apply the results to the patients? (จะประยุกต์ใช้ผลการศึกษากับผู้ป่วยอย่างไร)

Section D: Will the results help locally? (ผลการศึกษาจะช่วยท้องถิ่นหรือไม่)

10) Can the results be applied to your local population/ in your context? (จะสามารถประยุกต์ใช้ผลลัพธ์จากการศึกษากับประชาชนในท้องถิ่นหรือในบริบทของเราได้หรือไม่)

ข้อนี้เป็นการพิจารณาความตรงภายนอก (external validity หรือ generalizability) โดยผู้อ่านควรเริ่มพิจารณาว่าผู้ป่วยที่เราจะนำผลการศึกษาไปประยุกต์ใช้นั้นตรงกับเกณฑ์คัดเข้า (inclusion criteria) และไม่ได้อยู่ในเกณฑ์ในการคัดออก (exclusion criteria) หรือไม่ จากนั้นจึงพิจารณาลักษณะพื้นฐานของผู้เข้าร่วมการศึกษา (baseline characteristics) ว่าคล้ายคลึงกับลักษณะของผู้ป่วยที่เราจะนำผลการศึกษาไปประยุกต์ใช้หรือไม่

นอกเหนือจากการพิจารณาความตรงภายนอกแล้ว ในข้อ 10 นี้ยังต้องพิจารณาถึงผลลัพธ์ที่วัดว่าครอบคลุมผลลัพธ์ที่สำคัญทางคลินิกหรือไม่ ซึ่งในแคลสป์ฉบับก่อนหน้าจะแยกเป็นข้อคำถามเฉพาะว่า “Were all clinically important outcomes considered? (มีการพิจารณาผลลัพธ์ที่มีความสำคัญทางคลินิกไว้อย่างครบถ้วนหรือไม่)”

แต่ในแคลสปีฉบับปรับปรุงได้นำข้อคำถามดังกล่าวมา
รวมกับข้อที่ 10 แทน โดยแนวทางพิจารณาความ
ครอบคลุมผลลัพธ์ทางคลินิกที่สำคัญ ผู้อ่านจะทราบได้
จากการทบทวนข้อมูลเรื่องโรคและยาที่เกี่ยวข้อง โดย
หลักการพิจารณาคลายกับการทำ SOAP (Subjective
data-Objective data-Assessment-Plan) ในส่วน
ของการติดตามประสิทธิภาพและความปลอดภัย
นั่นเอง

11) Would the experimental intervention
provide greater value to the people in
your care than any of the existing
intervention? (สิ่งแทรกแซงที่ศึกษานั้นมีคุณค่า
ต่อผู้ป่วยที่เราากำลังดูแลมากกว่าสิ่งแทรกแซงใดๆ
ที่มีอยู่แล้วหรือไม่)

คำถามข้อนี้ถูกเพิ่มเข้ามาในแคลสปีฉบับ
ปรับปรุง ซึ่งจะให้พิจารณาว่าหากจะนำสิ่งแทรกแซง
ในการศึกษานั้นๆ ไปใช้ในบริบทของเรา เช่น
ในโรงพยาบาลที่ปฏิบัติงาน จะต้องลงทุนเรื่องเวลา
ค่าใช้จ่าย ตลอดจนจัดการอบรมหรือพัฒนาทักษะใหม่ๆ
ให้แก่บุคลากรในโรงพยาบาล เพื่อให้สามารถใช้
สิ่งแทรกแซงดังกล่าวได้อย่างถูกต้องก่อนหรือไม่
นอกจากนี้อาจพิจารณาว่าเราสามารถยุติการลงทุน
ในสิ่งแทรกแซงเดิมที่มีใช้มาก่อนแล้ว เพื่อมาลงทุนกับ
สิ่งแทรกแซงใหม่นี้หรือไม่

หลุมพรางการใช้งานเครื่องมือแคลสปีเพื่อ ประเมินความน่าเชื่อถือของงานวิจัยรูปแบบ RCT

จากประสบการณ์การสอนและการสังเกต
สิ่งที่นิตเภสัชศาสตร์นำเสนอเกี่ยวกับการประเมิน
ความน่าเชื่อถือของงานวิจัยรูปแบบ RCT โดยใช้
เครื่องมือแคลสปี มีประเด็นที่พึงระวังสรุปได้ ดังต่อไปนี้

1. การประเมินการสุ่มในงานวิจัยรูปแบบ
RCT โดยพิจารณาจากคำว่า randomization

เพียงอย่างเดียวไม่เพียงพอที่จะสรุปว่างานวิจัยนั้น
มีการสุ่มที่เหมาะสม

การประเมินประเด็นดังกล่าวผู้อ่านควรระบุ
วิธีแบ่งกลุ่มอย่างสุ่ม (random allocation) และวิธี
ปกปิดผลการแบ่งกลุ่ม (allocation concealment)¹
ให้ได้เสียก่อน จึงจะประเมินได้ว่างานวิจัยนั้นๆ
มีกระบวนการสุ่มที่เหมาะสมหรือไม่

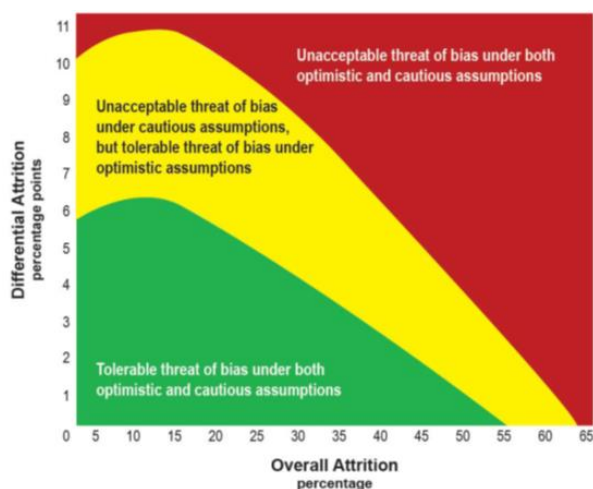
2. การประเมินจำนวนผู้ออกจากการศึกษา
ก่อนกำหนดที่อาจทำให้เกิดอคติต่อผลการศึกษา

เพื่อให้ชัดเจนในประเด็นดังกล่าว หน่วยงาน
What Works Clearinghouse (WWC) ของ
สหรัฐอเมริกาได้วางแนวทางการพิจารณาสัดส่วนของ
ผู้ออกจากการศึกษาที่กำหนดต่อความเสี่ยงในการ
เกิดอคติกับผลการศึกษาดังแสดงในรูปที่ 3⁴

จากรูปที่ 3 หากพิจารณาความแตกต่างของ
ผู้ที่ออกจากการศึกษาระหว่างกลุ่มและจำนวนผู้
ออกจากการศึกษาที่กำหนดทั้งหมดแล้ว พบว่า ตกอยู่
ในพื้นที่กราฟสีแดงหรือสีเขียว แสดงว่ามีโอกาสเกิด
อคติจากการคัดเลือกกลุ่มตัวอย่าง (selection bias)
สูงหรือต่ำตามลำดับ แต่ถ้าตกในพื้นที่สีเหลือง แสดงว่า
ผลการศึกษามีโอกาสเกิดอคติในระดับปานกลาง
จะต้องพิจารณาจากข้ออื่นๆ ร่วมด้วย โดยเฉพาะ
อย่างยิ่งความคล้ายคลึงกันของลักษณะพื้นฐานของ
ผู้เข้าร่วมการศึกษาระหว่างกลุ่ม จึงจะสรุปได้ว่ามี
โอกาสเกิดอคติต่อผลการศึกษามากน้อยเพียงใด ทั้งนี้
มีข้อสังเกตว่าสัดส่วนของการออกจากการศึกษา
ก่อนกำหนดที่แตกต่างกันระหว่างกลุ่ม จะส่งผลต่อ
อคติกับผลการศึกษามากกว่าสัดส่วนของผู้ที่ออกจาก
การศึกษาก่อนกำหนดทั้งหมด

ตัวอย่างเช่น การศึกษา A เปรียบเทียบกับ
ยาหลอก เมื่อสิ้นสุดการศึกษาพบว่ากลุ่มที่ได้ยา A
และยาหลอกมีคนออกจากการศึกษาก่อนกำหนด

คิดเป็น 12% และ 20% ของจำนวนที่ถูกสุ่มมา ตั้งแต่แรกตามลำดับ เมื่อคำนวณจำนวนคนที่ ออกจากการศึกษาก่อนกำหนดทั้งหมดจะเท่ากับ $(12\% + 20\%) / 2 = 16\%$ โดยความแตกต่างระหว่าง กลุ่มคิดเป็น $20\% - 12\% = 8\%$ ดังนั้นจะตกบนพื้นที่ กราฟสีเหลือง จึงมีความเสี่ยงต่อการเกิดอคติในระดับ ปานกลาง ต้องพิจารณาข้อคำถามอื่นร่วมด้วย เป็นต้น อย่างไรก็ตามนอกจากจะพิจารณาพื้นที่ในกราฟ ดังกล่าวแล้ว การมีคนออกจากการศึกษาก่อนกำหนด เป็นจำนวนมาก (เช่น มากกว่าร้อยละ 20) ก็อาจเป็น สัญญาณของการเกิดอคติจากการคัดเลือกกลุ่มได้ เช่นกัน เนื่องจากการมีคนออกไปเป็นจำนวนมาก อาจทำให้ความคล้ายคลึงกันของผู้เข้าร่วมการศึกษาระหว่างกลุ่ม (ซึ่งเป็นผลมาจากการสุ่มที่เหมาะสม) หดสั้นไป



รูปที่ 3 ความสัมพันธ์ระหว่างเปอร์เซ็นต์การออกจากการศึกษาก่อนกำหนดกับความเสี่ยงต่อการเกิดอคติ กับผลการศึกษา (ที่มา: What Works Clearinghouse, Standard Handbook Version 4.0⁴⁾)

3. การระบุว่าเป็น single-, double- หรือ triple-blind โดยไม่มีรายละเอียดอื่นๆ ไม่เพียงพอที่จะ ประเมินคำถามข้อที่ 4 (Were the participants,

the investigators, the people assessing/ analysing outcomes' blinded?) ว่า 'Yes'

ผู้วิจัยควรระบุรายละเอียดของการปกปิดไว้ อย่างชัดเจนว่าปกปิดใครบ้างและปกปิดอย่างไร เช่น “neither the person doing the assessment nor the study participant could identify the intervention being assessed.” อย่างไรก็ตามหาก ไม่ได้กล่าวถึงกลุ่มคนที่ถูกปกปิดไว้อย่างชัดเจน แต่ได้ ระบุว่าใช้ active placebo, identical placebo หรือ dummy ก็จะสามารถประเมินได้เช่นกันว่ามีการ ปกปิดแบบสองทางแล้ว⁵

4. การประเมินความคล้ายคลึงกันของ ลักษณะพื้นฐานของผู้เข้าร่วมการศึกษาระหว่างกลุ่ม ทดลองและกลุ่มควบคุม

ประเด็นที่มักถูกมองข้ามในข้อนี้ คือ การพิจารณาว่าลักษณะพื้นฐานที่รายงานไว้ ครอบคลุม ลักษณะสำคัญที่อาจส่งผลต่อผลลัพธ์ในการศึกษา หรือไม่ ซึ่งเป็นสิ่งที่ได้จากการทบทวนความรู้เรื่องโรค และความเชี่ยวชาญของผู้อ่านงานวิจัย

ตัวอย่างเช่น ปัจจัยสำคัญที่สัมพันธ์กับการ เสียชีวิตในผู้ป่วยโรคหัวใจล้มเหลว (heart failure with reduced ejection fraction; HFrEF) ได้แก่ เพศ อายุ ดัชนีมวลกาย การมีโรคร่วม โดยเฉพาะอย่างยิ่งโรคกล้ามเนื้อหัวใจขาดเลือด โรคเบาหวาน โรคความดันโลหิตสูง และการได้รับยาบางกลุ่มที่ช่วยลดอัตราการเสียชีวิตในผู้ป่วยกลุ่มนี้ นี้ เช่น beta-blockers, angiotensin converting enzyme inhibitors (ACEIs), angiotensin receptor blockers (ARBs), mineralocorticoid receptor antagonists และ angiotensin receptor blockers + neprilysin inhibitors (ARNi) เป็นต้น ดังนั้น หากในการศึกษาผลของยาต่อการเสียชีวิตในผู้ป่วย โรคหัวใจล้มเหลวชนิด HFrEF

ที่ผู้วิจัยไม่ได้แสดงลักษณะเหล่านี้ไว้ตั้งแต่ต้น ผู้อ่านควรตระหนักว่าผลลัพธ์ของการศึกษาอาจถูกรบกวนจากปัจจัยเหล่านี้ โดยเฉพาะอย่างยิ่งหากไม่มีกระบวนการสุ่มที่เหมาะสม

5. การประเมินความเท่าเทียมกันของการรักษาและสิ่งแทรกแซงร่วมที่ให้ในระหว่างการศึกษา

ประเด็นดังกล่าวมักพบว่าผู้อ่านงานวิจัยจะพิจารณาจากความคล้ายคลึงกันของลักษณะพื้นฐานระหว่างกลุ่ม ซึ่งเป็นสิ่งที่ไม่ถูกต้อง เนื่องจากลักษณะพื้นฐานเป็นค่าที่วัดหรือเก็บข้อมูล ณ ขณะที่สุ่มเพียงจุดเวลาเดียว ในขณะที่สิ่งแทรกแซงร่วมเป็นสิ่งที่ผู้เข้าร่วมการศึกษามีโอกาสได้รับไปตลอดจนการศึกษาเสร็จสิ้น ดังนั้น การมีลักษณะพื้นฐานที่คล้ายคลึงกัน จึงไม่ใช่เครื่องรับประกันว่าทั้งสองกลุ่มจะได้รับสิ่งแทรกแซงร่วมเหมือนกัน

ตัวอย่างเช่น ในการศึกษาประสิทธิผลของยา A เปรียบเทียบกับยา B ต่อการป้องกันการติดเชื้อโควิด-19 การควบคุมสิ่งแทรกแซงร่วมอาจเป็นการขอให้ผู้เข้าร่วมการศึกษาทั้งสองกลุ่มปฏิบัติตามแนวทางการป้องกันการติดเชื้อที่เหมือนกัน เช่น การหมั่นล้างมือ การรักษาระยะห่างทางสังคม การสวมหน้ากากอนามัย เป็นต้น หรือในการศึกษา RAPID trial ซึ่งเปรียบเทียบขนาดของยาเฮพาริน (heparin) ที่ใช้ในผู้ป่วยโควิด-19 ต่อการลดความรุนแรงของโรค ผู้วิจัยได้แสดงให้เห็นว่าตลอดระยะเวลาการติดตาม ผู้ป่วยจะได้รับยา รวม ได้แก่ systemic corticosteroid, remdesivir และ tocilizumab ในสัดส่วนที่ใกล้เคียงกันระหว่างกลุ่ม⁶ ซึ่งเป็นสิ่งยืนยันว่าผู้ป่วยทั้งสองกลุ่มได้รับการรักษาที่เท่าเทียมกันนั่นเอง

6. ประเด็นเกี่ยวกับเกณฑ์คัดเข้า (inclusion criteria) และเกณฑ์คัดออก (exclusion criteria)

สิ่งที่พบว่ามีคนเข้าใจผิดคือ ผู้อ่านมักเข้าใจว่าเกณฑ์คัดออกคือสิ่งตรงข้ามกับเกณฑ์คัดเข้า ซึ่งไม่ถูกต้อง เพราะโดยหลักการแล้วผู้เข้าร่วมงานวิจัยจะถูกคัดเลือกเข้ามาในการศึกษาก่อนตามเกณฑ์คัดเข้า แต่หากพบว่าคนผู้นั้นมีสถานะบางอย่างที่ตรงกับเกณฑ์คัดออก ซึ่งมักเป็นสถานะที่ส่งผลต่อความปลอดภัยหรือความร่วมมือในการใช้ยา คนผู้นั้นจะถูกคัดออกจากการศึกษา กล่าวอีกนัยหนึ่งว่าผู้เข้าร่วมการศึกษาคนหนึ่งที่ผ่านเกณฑ์คัดเข้าจะมีโอกาสถูกคัดออกได้หากมีสถานะหรือเงื่อนไขตามเกณฑ์คัดออกที่กำหนด ตัวอย่างเช่น เกณฑ์คัดเข้าคือ ผู้ที่มีอายุ 18 ปีขึ้นไป เกณฑ์คัดออกคือ ผู้ที่กำลังตั้งครรภ์หรือให้นมบุตร ผู้ที่มีความผิดปกติของตับหรือไตรุนแรง หรือผู้ที่อาการของโรคประจำตัวยังไม่คงที่เป็นต้น

7. การนำผลการวิเคราะห์กลุ่มย่อย (subgroup analysis) ไปใช้กับผู้ป่วยโดยตรง

ผลการวิเคราะห์กลุ่มย่อยอาจพบนัยสำคัญทางสถิติโดยบังเอิญ ซึ่งเป็นผลของความคลาดเคลื่อนชนิดที่หนึ่ง (type I error) จากการวิเคราะห์ทางสถิติซ้ำๆ (multiple comparisons) ตัวอย่างเช่น การศึกษา ISIS-2 ซึ่งศึกษาผลของ streptokinase และแอสไพรินต่อการเกิดโรคกล้ามเนื้อหัวใจขาดเลือด ผลการวิเคราะห์กลุ่มย่อยแสดงให้เห็นว่าประสิทธิผลของ streptokinase และแอสไพรินสัมพันธ์กับลักษณะการเกิดอย่างมีนัยสำคัญทางสถิติ โดยในการศึกษา ISIS-2 ดังกล่าวผู้วิจัยต้องการแสดงให้เห็นว่านัยสำคัญทางสถิติที่พบนั้นเกิดขึ้นจากความบังเอิญ⁷

ทั้งนี้การวิเคราะห์กลุ่มย่อยควรทำเพื่อสำรวจความสัมพันธ์ ซึ่งจะนำไปสู่การตั้งสมมติฐานสำหรับออกแบบการศึกษาในอนาคต แต่ไม่ควรนำผลของการวิเคราะห์กลุ่มย่อยมาใช้กับผู้ป่วยโดยตรง นอกจากจะมีเหตุผลสนับสนุนที่ดีเพียงพอ เช่น

การวิเคราะห์กลุ่มย่อยนั้นได้ถูกวางแผนไว้ล่วงหน้า ซึ่งผู้อ่านสามารถทราบได้จากการอ่านโพรโทคอล (protocol) ของงานวิจัย และมีการปรับผลจากการวิเคราะห์ซ้ำด้วยวิธีทางสถิติแล้ว เป็นต้น

เครื่องมืออื่นๆ ที่ใช้ประเมินความน่าเชื่อถือ ของงานวิจัยรูปแบบ RCT

ดังที่ได้กล่าวไปแล้วว่าเครื่องมือที่ใช้ประเมินความน่าเชื่อถือของงานวิจัยรูปแบบ RCT มีหลายชนิด หากพิจารณาเครื่องมือที่ใช้อย่างแพร่หลาย สามารถแบ่งได้เป็น 3 กลุ่มตามลักษณะการใช้งาน ได้แก่

1. เครื่องมือที่ใช้ประเมินวรรณกรรมเพื่อนำไปใช้กับผู้ป่วย ได้แก่ CASP², FRISBE(E)⁸⁻⁹ และ GATE (Graphic Appraisal Tools for Epidemiology)¹⁰

ในกลุ่มนี้จะเป็นเครื่องมือที่ใช้ประเมินความตรงภายในของผลการศึกษา ประเมินขนาดและความแม่นยำของผลลัพธ์ ตลอดจนประเมินการนำไปใช้กับผู้ป่วย เครื่องมือในกลุ่มนี้จึงเหมาะกับการประเมินงานวิจัยเพื่อนำไปใช้ดูแลผู้ป่วยแต่ละราย โดยเครื่องมือ FRISBE(E) ย่อมาจาก Follow-up – Randomisation – Intention-to-treat analysis – Similarity – Blinding – Equal treatment – (Equivalent to my patient)⁸⁻⁹ ซึ่งคล้ายกับแคสพิ แต่จดจำง่ายกว่า อย่างไรก็ตามมีข้อจำกัดที่ไม่ได้ประเมินความชัดเจนของคำถามงานวิจัย จึงอาจทำให้ผู้ประเมินมองข้ามประเด็นดังกล่าวได้ สำหรับ GATE เป็นเครื่องมือที่มีมิติในการประเมินความน่าเชื่อถือครอบคลุมทุกประเด็นความตรงภายในที่ควรพิจารณาสามารถปรับใช้กับการศึกษาได้หลากหลายรูปแบบ และเห็นภาพรวมของการศึกษาค่อนข้างชัดเจน แต่เครื่องมือดังกล่าวมีรายละเอียดให้พิจารณามาก ทำให้ยากแก่การจดจำและนำไปใช้¹⁰

2. เครื่องมือที่ใช้ประเมินวรรณกรรมเพื่อการสังเคราะห์ข้อมูลใหม่ เช่น การวิเคราะห์ห่อภิมาณ (meta-analysis) ได้แก่ Cochrane risk of bias (RoB)¹¹ และ JADAD⁵

ในกลุ่มนี้จะเป็นเครื่องมือที่เน้นการประเมินความตรงภายในการศึกษาและมีระบบการให้คะแนนที่ชัดเจน เช่น คะแนน 0-5 สำหรับ JADAD และ unclear/ low/ high risk of bias ใน RoB ผู้อ่านจะสังเกตว่าเครื่องมือในกลุ่มนี้ไม่ได้พิจารณาเรื่องขนาดผลลัพธ์และการนำไปใช้กับผู้ป่วยเหมือนเครื่องมือกลุ่มแรก แต่เป็นการประเมินเพื่อเป็นส่วนหนึ่งที่ช่วยพิจารณาว่าผลการวิเคราะห์ห่อภิมาณน่าเชื่อถือมากน้อยเพียงใด กล่าวคือ ถ้างานวิจัยที่นำมารวมกันล้วนแต่เป็นงานวิจัยที่มีคุณภาพดี ก็จะเป็นส่วนหนึ่งที่ทำให้ผลการวิเคราะห์ห่อภิมาณมีความน่าเชื่อถือตามไปด้วย

3. เครื่องมือที่ใช้เป็นแนวทางการเขียนรายงานวิจัยรูปแบบ RCT ได้แก่ Consolidated Standards of Reporting Trials (CONSORT)¹²

เครื่องมือกลุ่มนี้ไม่ได้มีวัตถุประสงค์หลักเพื่อใช้ประเมินความน่าเชื่อถือของงานวิจัย แต่ก็มีผู้ประยุกต์ใช้แนวทางดังกล่าวมาช่วยประเมินว่างานวิจัยนั้นน่าเชื่อถือหรือไม่ ผู้อ่านอาจตั้งข้อสังเกตว่าการนำเครื่องมือดังกล่าวมาใช้ประเมินความน่าเชื่อถือของงานวิจัยนั้นทำได้ยาก เนื่องจากเป็นแนวทางกลางๆ ที่ไม่ได้มีรายละเอียดว่าสิ่งที่ถูกรายงานไว้นั้นเหมาะสมหรือนำไปสู่ข้อคิดต่อผลลัพธ์การศึกษาหรือไม่

ตัวอย่างเช่น กระบวนการสุ่มใน CONSORT แนะนำเพียงว่าผู้วิจัยควรรายงานวิธีสร้างลำดับการสุ่ม (sequence generation) และวิธีปกปิดผลการแบ่งกลุ่ม (allocation concealment) แต่ไม่ได้กำหนดว่าวิธีที่เหมาะสมเป็นอย่างไร หรือการกำหนดว่าผู้วิจัยควรรายงานจำนวนผู้ที่ออกจากการศึกษาก่อนกำหนด

แต่ไม่ได้ระบุว่าจำนวนคนที่ออกจากการศึกษาก่อนกำหนดมากเท่าใดที่จะทำให้เกิดอคติต่อผลการศึกษาเป็นต้น นอกจากนี้ CONSORT ยังไม่ครอบคลุมการพิจารณาผลลัพธ์ของการศึกษา โดยระบุเพียงว่าผู้วิจัยควรรายงานผลลัพธ์อะไรบ้าง เช่น ขนาดผลลัพธ์และช่วงความเชื่อมั่น เป็นต้น รวมทั้งไม่ครอบคลุมประเด็นการนำไปใช้กับผู้ป่วย รายละเอียดของเครื่องมือต่างๆ ที่ใช้ประเมิน RCT (ตารางที่ 3)

บทสรุป

แอสพีเป็นหนึ่งในเครื่องมือที่ใช้ประเมินความน่าเชื่อถือของงานวิจัยรูปแบบสุ่มที่มีการควบคุม ในพ.ศ.2563 มีการปรับปรุงเครื่องมือดังกล่าวโดยการประเมินความตรงภายในของผลการศึกษายังคงไว้เช่นเดิม แต่ได้เพิ่มการพิจารณาวิธีการได้มาซึ่งผลลัพธ์นั้นๆ เช่น การคำนวณขนาดตัวอย่าง การเลือกสถิติทดสอบ เป็นต้น รวมทั้งมีการเพิ่มมุมมองเชิงบริหารจัดการเพื่อประเมินว่าสิ่งแทรกแซงที่จะนำมาใช้นั้น

สามารถทดแทนสิ่งแทรกแซงมาตรฐานที่ใช้อยู่ในปัจจุบันหรือไม่ แม้การเพิ่มข้อพิจารณาดังกล่าวจะทำให้ประเมินได้ยากขึ้น แต่ก็ทำให้ประเมินได้รอบด้านมากกว่าการพิจารณาด้วยมุมมองทางคลินิกเพียงอย่างเดียว ทั้งนี้ ประเด็นที่พึงระวังเกี่ยวกับการใช้เครื่องมือแอสพี สามารถหลีกเลี่ยงได้ด้วยการฝึกใช้เครื่องมือบ่อยๆ ร่วมกับการตั้งข้อสงสัยว่าประเด็นที่กำลังพิจารณาอยู่นั้น ประเมินไปเพื่ออะไร

อนึ่ง ยังมีเครื่องมืออื่นๆ หลายชนิดที่ใช้ประเมินความน่าเชื่อถือของงานวิจัยได้ แต่ไม่ว่าจะใช้เครื่องมือใดก็ตาม สุดท้ายแล้วผู้อ่านจะต้องสรุปให้ได้ว่า ผลการศึกษามีความถูกต้องน่าเชื่อถือหรือไม่ ผลการศึกษาเป็นอย่างไร และจะประยุกต์ใช้ผลการศึกษาดังกล่าวกับผู้ป่วยที่เราดูแลอย่างไร ทั้งนี้ เพื่อเป้าหมายในการดูแลผู้ป่วยอย่างมีประสิทธิภาพบนพื้นฐานของหลักฐานเชิงประจักษ์ทางการแพทย์นั่นเอง

ตารางที่ 3 การเปรียบเทียบเครื่องมือต่างๆ ที่ใช้ประเมินงานวิจัยรูปแบบ RCT

ประเด็นที่ประเมิน	CASP ²	FRISBE(E) ⁸⁻⁹	GATE ¹⁰	Cochrane RoB ¹¹	JADAD ⁵
ผลการศึกษามีความถูกต้องหรือไม่					
อคติจากการเลือกกลุ่มตัวอย่าง (selection bias)	2 Randomisation 3 Attrition/ ITT/ FU 5 Similarity	2 Randomisation (R) 1 FU (F) 3 ITT (I) 4 Similarity (S)	- Allocation (A) - Maintenance in allocated groups (M) (i.e., completeness of FU, compliance, contamination)	- Random sequence generation - Allocation concealment - Incomplete outcome data	1 Randomisation 3 Attrition (explanation)
อคติเชิงข้อมูล (information bias)	4 Blinding	5 Blinding (B)	Blinding (B) Objective outcome (O)	- Blinding of participants - Blinding of outcome assessment	2 Double blinding
ปัจจัยกวน (confounders)	5 Similarity 6 Co-intervention	6 Equal treatment (E)	Maintenance in allocated groups (co-intervention) (M) 1 Study design (AMBOM) 2 Study ANALyses (AN)	- Other bias	◎

ตารางที่ 3 (ต่อ)

ประเด็นที่ประเมิน	CASP ²	FRISBE(E) ⁸⁻⁹	GATE ¹⁰	Cochrane RoB ¹¹	JADAD ⁵
การพิจารณาอื่นๆ	1 PICO (research question)	◎	- PECOT - Recruitment appropriate (R) - Mean/ Median FU time - Validity of the analysis - Selective reporting	- Selective reporting - Other bias	◎
ผลการศึกษาเป็นอย่างไร	- Comprehensively report treatment effect? - Report precision (confidence interval)?	- How large? - How precise?	3 Study numbers reported – Random error 4 Study effect estimates reported	◎	◎
ประโยชน์ อันตราย และต้นทุน	◎	◎	◎	◎	◎
จะประยุกต์ใช้ผลการศึกษากับผู้ป่วยอย่างไร					
การประยุกต์ใช้กับผู้ป่วย	◎	7 Equivalence to my patient (E)	5 Generalisability (R)	◎	◎
รายงานผลลัพธ์ที่สำคัญไว้ครบถ้วน	◎ [‡]	◎	◎	◎	◎
มีคุณค่ามากกว่าสิ่งที่ใช้ในปัจจุบัน	◎	◎	◎	◎	◎

ตารางที่ 3 (ต่อ)

ประเด็นที่ประเมิน	CASP ²	FRISBE(E) ⁸⁻⁹	GATE ¹⁰	Cochrane RoB ¹¹	JADAD ⁵
จุดแข็งของเครื่องมือ	ข้อคำถามน้อยข้อและครอบคลุมประเด็นสำคัญ	ชื่อเครื่องมือเป็นอักษรย่อ ทำให้ง่ายต่อการจดจำ	ปรับใช้กับการศึกษาได้ หลากหลายรูปแบบ และครอบคลุมทุกประเด็นความตรงภายในที่ควรพิจารณา	ครอบคลุมประเด็นความตรงภายในไว้อย่างละเอียดครบถ้วน	ประเมินง่ายและมีระบบการให้คะแนนที่ชัดเจนที่สุด
ข้อจำกัดที่สำคัญ	ชื่อเครื่องมือไม่สื่อถึงสิ่งที่ประเมิน ทำให้ยากต่อการจดจำ นอกจากนี้ในฉบับปรับปรุงได้เพิ่มการพิจารณามุมมองเชิงบริหารจัดการ ทำให้ประเมินได้ยากขึ้น	ไม่ได้ประเมินความชัดเจนของคำถามงานวิจัย	มีรายละเอียดให้พิจารณา มาก ทำให้ยากต่อการนำไปใช้	ประเมินยาก เนื่องจากมีข้อ other risk of bias ซึ่งขึ้นกับประสบการณ์และความเชี่ยวชาญของผู้อ่านงานวิจัย	ไม่ครอบคลุมประเด็นสำคัญ เช่น allocation concealment, similarity และขาดความยืดหยุ่น เช่น การศึกษาที่ไม่ได้ทำ double-blinding แต่ปกปิดผู้ประเมินผลลัพธ์และวัดผลลัพธ์ที่เป็นปรวิสัย แต่ก็ยังถูกประเมินเป็น 0 ในส่วนดังกล่าว

REFERENCES

1. Na-Ek N. The guide to critically appraise a randomised controlled trials (RCT) using a CASP tool. Chiangrai Medical Journal. 2020;12(2):131–55.
2. Critical Appraisal Skills Programme. 2020-CASP checklists. [Internet]. 2020 [cited 2021 Sep 25]. Available from: <https://casp-uk.net/casp-tools-checklists/>.
3. Kirkwood BR, Sterne JAC. Essential medical statistics. 2nd ed. Malden, Massachusetts: Blackwell Science; 2003.
4. What Works Clearinghouse. What Works Clearinghouse™ Standards Handbook (Version 4.0) [Internet]. [cited 2020 Aug 10]. Available from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf
5. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17(1):1-12.
6. Sholzberg M, Tang GH, Rahhal H, AlHamzah M, Kreuziger LB, Áinle FN, et al. Effectiveness of therapeutic heparin versus prophylactic heparin on death, mechanical ventilation, or intensive care unit admission in moderately ill patients with covid-19 admitted to hospital: RAPID randomised clinical trial. BMJ. 2021;375:n2400.
7. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17187 cases of suspected acute myocardial infarction: ISIS-2. Lancet. 1988;332(8607):349–60.
8. Xiong GL, Adams J. FRISBEE: does the study fly in the face of evidence-based medicine? Curr Psychiatr. 2007;6(12):96.
9. Duke University Medical Center Library & Archives. Evidence-based practice: appraise [Internet]. [cited 2020 Aug 11]. Available from: <https://guides.mclibrary.duke.edu/ebm/appraise>
10. Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, et al. The GATE frame: critical appraisal with pictures. Evid Based Med. 2006;11(2):35-8
11. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
12. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012;308(24):2594–604.



โรงพยาบาลเชียงรายประชานุเคราะห์
Chiangrai Prochanukroh Hospital

ISSN 1906-649X ปีที่ 13 ฉบับที่ 3 เดือน ธันวาคม 2564

เชียงรายเวชสาร

วารสารวิชาการ โรงพยาบาลเชียงรายประชานุเคราะห์ | Chiangrai Medical Journal