

แนวทางการประเมินงานวิจัยนิตการทดลองแบบสุ่ม ที่มีการควบคุมด้วยเครื่องมือแคสพี

ณัฐ นาเอก ภ.บ. (บริหารเภสัชกรรม), MSc, PhD (Clinical Epidemiology)*

วัตถุประสงค์

1. เพื่อให้ผู้อ่านทราบแนวทางการประเมินงานวิจัยปฐมภูมิรูปแบบสุ่มที่มีการควบคุมด้วยเครื่องมือแคสพี
2. เพื่อให้ผู้อ่านสามารถนำเครื่องมือแคสพีไปประยุกต์ใช้ในการประเมินงานวิจัยปฐมภูมิรูปแบบสุ่มที่มีการควบคุมได้

บทคัดย่อ

งานวิจัยในรูปแบบสุ่มที่มีการควบคุม (randomised controlled trial; RCT) แม้จะมีความน่าเชื่อถือสูงสุดในระดับของงานวิจัยปฐมภูมิ (primary literature) แต่การไม่มีระเบียบวิธีวิจัยที่ดี อาจทำให้ผลการวิจัยผิดพลาดอันเนื่องมาจากอคติที่แฝงอยู่ได้ ดังนั้นการประเมินความน่าเชื่อถือของงานวิจัย RCT จึงเป็นทักษะพื้นฐานที่บุคลากรทางการแพทย์ทุกคนพึงมี ปัจจุบันนี้มีเครื่องมือที่ใช้เพื่อประเมินความน่าเชื่อถือของ RCT หลายชนิด โดยเครื่องมือแต่ละชนิดก็มีจุดแข็งและข้อจำกัดที่แตกต่างกันไป ในบทความฉบับนี้จะแนะนำเครื่องมือแคสพี (critical appraisal skill programme ; CASP) เนื่องจากมีข้อคำถามเพียง 11 ข้อจึงง่ายต่อการนำไปประยุกต์ใช้ นอกจากนี้เครื่องมือยังครอบคลุมประเด็นคำถามสำคัญไว้อย่างครบถ้วน แต่ทั้งนี้ไม่ว่าจะเลือกใช้เครื่องมือใดก็ตาม ผู้อ่านจะต้องสรุปคำตอบของข้อคำถามที่สำคัญสามข้อให้ได้ ซึ่งได้แก่ 1) ผลลัพธ์การศึกษาที่มีความตรงหรือไม่ 2) ผลลัพธ์จากการศึกษาเป็นอย่างไร และ 3) จะประยุกต์ใช้ผลลัพธ์ดังกล่าวกับผู้ป่วยได้อย่างไร (หรือผลการศึกษจะช่วยท้องถิ่นหรือไม่) ทั้งนี้เพื่อเป้าหมายสำคัญ คือการเลือกใช้หลักฐานเชิงประจักษ์ที่มีความถูกต้องและประยุกต์ใช้กับผู้ป่วยได้อย่างเหมาะสม อันจะเป็นการคงไว้ซึ่งประโยชน์สูงสุดที่จะเกิดขึ้นกับตัวผู้ป่วยนั่นเอง

คำสำคัญ : การวิจัยแบบสุ่มที่มีการควบคุม การประเมินความน่าเชื่อถือ เครื่องมือแคสพี

*สาขาวิชาบริหารเภสัชกรรม คณะเภสัชศาสตร์ มหาวิทยาลัยพะเยา

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

วันที่รับเรื่อง : 25 พฤษภาคม 2563 วันที่ส่งแก้ไข : 1 กรกฎาคม 2563 วันที่ตีพิมพ์ : 27 กรกฎาคม 2563

THE GUIDE TO CRITICALLY APPRAISE A RANDOMISED CONTROLLED TRIAL (RCT) USING A CASP TOOL

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

ABSTRACT

Although a randomised controlled trial (RCT) is universally accepted as the most trustworthy study designs of primary literature, without a vigorous methodology, its findings can be biased and misleading. Therefore, critical appraisal is an essential skill for all health care professionals. Currently, there are several tools available for critically appraising RCT with their unique strengths and limitations. In this article, an appraisal tool called 'critical appraisal skill programme' (CASP) is introduced since it contains only 11 questions, making it easy to adopt. Besides, the tool covers the main issues in criticising RCT. Regardless of a chosen device, readers should ultimately be able to answer three following critical questions: 1) Are the results valid? 2) What are the results? and 3) How can the results be applied to the patients (or will the results help locally)? The goal is for clinicians to endorse the right evidence and undertake proper implementation to maximise the benefit for patients.

KEYWORDS

Randomised controlled trial (RCT), critical appraisal, critical appraisal skill programme (CASP)

*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 : 25 May 2020 Revise date : 1 July 2020 Publish date : 27 July 2020

บทนำ

ในแต่ละวันงานวิจัยทางคลินิกถูกตีพิมพ์ออกมาเผยแพร่สู่สาธารณชนเป็นจำนวนมาก และงานวิจัยรูปแบบสุ่มที่มีการควบคุม (RCT) เป็นรูปแบบงานวิจัยปฐมภูมิ (primary literature) ที่มีความน่าเชื่อถือสูงสุด ทำให้มีการนำผลลัพธ์จากงานวิจัยเหล่านี้มาประยุกต์ใช้ในการดูแลผู้ป่วยอย่างแพร่หลาย รวมถึงนำไปกำหนดเป็นมาตรฐานในการดูแลผู้ป่วย อย่างไรก็ตามหากผู้วิจัยไม่ได้ดำเนินการวิจัยตามระเบียบวิธีวิจัยที่ดี การศึกษาชนิด RCT นี้ ก็อาจเต็มไปด้วยอคติจนนำไปสู่ผลการศึกษาที่ไม่ถูกต้องได้ ดังนั้นการประเมินความน่าเชื่อถือของงานวิจัยชนิด RCT จึงเป็นทักษะพื้นฐานสำคัญที่บุคลากรทางการแพทย์พึงมี ทั้งนี้การประเมินงานวิจัยทางการแพทย์มีเป้าหมายสำคัญเพื่อตอบคำถาม 3 ข้อตามลำดับ ได้แก่ 1) ผลลัพธ์ของการศึกษามีความตรงหรือไม่ 2) ผลลัพธ์ของการศึกษาเป็นอย่างไร และ 3) จะประยุกต์ใช้ผลการศึกษานี้กับผู้ป่วยได้อย่างไร (หรือผลการศึกษาจะช่วยท้องถิ่นหรือไม่)

เครื่องมือในการประเมินงานวิจัยชนิด RCT นี้มีหลายเครื่องมือ เช่น Cochrane risk of bias¹, JADAD score², แนวทางการเขียนรายงานวิจัยรูปแบบ RCT (Consolidated Standards of Reporting Trials: CONSORT)³ โดยเครื่องมือแต่ละชนิดจะมีข้อดีข้อเสียที่แตกต่างกัน และไม่มีข้อกำหนดมาตรฐานว่าควรใช้เครื่องมือใดเป็นหลัก สำหรับบทความฉบับนี้ จะขอแนะนำการประเมินโดยใช้เครื่องมือแคสพี (Critical Appraisal Skill Programme; CASP) เนื่องจากเป็นเครื่องมือที่มีข้อคำถามเพียง 11 ข้อ ซึ่งครอบคลุมประเด็นสำคัญในการพิจารณาความน่าเชื่อถือของงานวิจัยรูปแบบ RCT ไว้อย่างครบถ้วน เครื่องมือแคสพีได้รับการคิดค้นและพัฒนาจากนักวิชาการจากมหาวิทยาลัยออกซ์ฟอร์ด

สหราชอาณาจักร ตั้งแต่ปี 1993 โดยได้พัฒนาขึ้นจากคำแนะนำของ Guyatt และคณะ^{4,5} ทั้งนี้ไม่ว่าผู้อ่านจะใช้เครื่องมือใดในการประเมินงานวิจัย สุดท้ายจะต้องสรุปคำตอบของคำถามสำคัญทั้ง 3 ข้อข้างต้น

บทความฉบับนี้ประกอบด้วยเนื้อหาสองส่วน เนื้อหาส่วนแรกจะเป็นการอธิบายขั้นตอนการประเมินงานวิจัยรูปแบบ RCT และเนื้อหาส่วนที่สองเป็นการแสดงตัวอย่างการประเมินงานวิจัย ซึ่งรายละเอียดจะแสดงไว้ในส่วนของภาคผนวก โดยในที่นี้จะยกงานวิจัย “Canagliflozin and cardiovascular and renal events in type 2 diabetes (CANVAS program trial)” โดย Bruce และคณะ⁶ ที่ตีพิมพ์ในวารสาร The New England Journal of Medicine ในปี พ.ศ. 2560 มาใช้เป็นตัวอย่งในการประเมิน

ขั้นตอนในการประเมินงานวิจัยตาม แนวทางของผู้เขียน

การประเมินงานวิจัยประกอบด้วย 4 ขั้นตอนที่สำคัญ ได้แก่ 1) สรุปรงานวิจัย 2) ทบทวนความรู้เรื่องโรคและข้อมูลยาที่เกี่ยวข้อง 3) ประเมินงานวิจัย และ 4) สรุปผลจากการประเมินงานวิจัย ทั้งนี้เพื่อให้ได้ข้อสรุปว่า ผลลัพธ์ของการศึกษามีความตรงหรือไม่ ผลลัพธ์ของการศึกษาเป็นอย่างไร และจะประยุกต์ใช้ผลลัพธ์นี้กับผู้ป่วยได้อย่างไร (หรือผลการศึกษาจะช่วยท้องถิ่นหรือไม่) โดยรายละเอียดในแต่ละขั้นตอนสามารถอธิบายได้ดังต่อไปนี้

1. สรุปงานวิจัย

เป็นการเขียนสรุปเนื้อหาจากบทความที่อ่าน ซึ่งเป็นสิ่งสำคัญที่ผู้อ่านควรทำเป็นอันดับแรก ทั้งนี้เพื่อเป็นการสรุปว่าบทความที่กำลังอ่านอยู่นั้นมีใจความสำคัญอย่างไร ดังนั้นควรเป็นเนื้อหาสรุปที่สั้น กระชับ (แนะนำว่าความยาวไม่ควรเกิน 2 หน้ากระดาษ A4) ทั้งนี้การเขียนสรุปบทความอาศัยหลักการของ PICO (Patient-Intervention - Comparator - Outcome) โดยควรพิจารณาข้อมูลจากภาคผนวก (หรือดัชนี) และ โพรโทคอลของการศึกษานั้นร่วมด้วยเสมอ

2. ทบทวนความรู้เรื่องโรคและข้อมูลยาที่เกี่ยวข้อง

ขอบเขตของการทบทวนความรู้เรื่องโรค จะพิจารณาจากโรคที่เกี่ยวข้องกับบทความ เช่น หากเป็นบทความที่เกี่ยวข้องกับการป้องกันปฐมภูมิ (primary prevention) อาจพิจารณาทบทวนข้อมูลที่เกี่ยวข้องกับสาเหตุของโรค ปัจจัยเสี่ยง ความชุกหรืออุบัติการณ์การเกิดโรคนั้น และแนวทางการวินิจฉัยโรคนั้นตามมาตรฐานการวินิจฉัย ณ เวลาที่ทำการศึกษานั้น แต่หากเป็นบทความที่เกี่ยวข้องกับการป้องกันทุติยภูมิ (secondary prevention) อาจพิจารณาข้อมูลการพยากรณ์ (prognosis) และการดำเนินไปของโรค ภาวะแทรกซ้อน และการรักษามาตรฐาน ณ ช่วงเวลาที่ดำเนินการศึกษานั้น (ไม่ใช่มาตรฐานการรักษา ณ ปัจจุบัน) นอกจากนี้ในการประเมินงานวิจัยที่ศึกษาประสิทธิผลของยา การทบทวนข้อมูลยาที่เกี่ยวข้องก็มีความสำคัญเช่นกัน

แหล่งข้อมูลที่สามารถใช้ทบทวนความรู้เรื่องโรค ได้แก่ วารสารตติยภูมิ (tertiary literature) เช่น ตำราทางการแพทย์ แนวทางการรักษาโรค บทความปริทัศน์วรรณกรรม (review literature) หรือฐานข้อมูลทางการแพทย์ออนไลน์ เช่น ฐานข้อมูล UpToDate® เป็นต้น

3. ประเมินงานวิจัยโดยใช้เครื่องมือแคสปี

เกณฑ์การประเมินแคสปีจะประกอบด้วย ข้อคำถามทั้งหมด 11 ข้อ แต่ละข้อจะให้ประเมินว่า ผู้วิจัย “ได้ทำ (Yes)” “ไม่ได้ทำ (No)” หรือ “ไม่สามารถระบุได้ (Can't tell)” โดยผู้อ่านสามารถเข้าถึงแบบประเมินแคสปีได้จากเว็บไซต์ <https://casp-uk.net/casp-tools-checklists/> ซึ่งคำถามทั้ง 11 ข้อนี้ จะถูกแบ่งเป็น 3 ส่วน ดังรายละเอียดต่อไปนี้

ส่วนที่หนึ่ง: ผลการศึกษามีความตรงหรือไม่ (Are the results valid?)

1. งานวิจัยได้ระบุถึงประเด็นที่ต้องการศึกษาไว้อย่างชัดเจนหรือไม่ (Did the trial address a clearly focused issue?)

ข้อนี้ผู้อ่านต้องพิจารณาว่า สามารถสกัดข้อมูล PICO ออกจากงานวิจัยได้หรือไม่ และเมื่อสกัดข้อมูลดังกล่าวออกมาแล้ว ผู้อ่านมองเห็นภาพรวมของประชากรที่ศึกษา สิ่งแทรกแซงที่ศึกษา สิ่งที่น่าสนใจเปรียบเทียบ รวมทั้งผลลัพธ์ในการศึกษานี้หรือไม่ โดยทั่วไปข้อมูลส่วนนี้สามารถประเมินได้ตั้งแต่การอ่านบทคัดย่อและส่วนท้ายของบทนำในงานวิจัย

2. มีการสุ่มให้สิ่งแทรกแซงแก่ผู้ป่วยหรือไม่ (Was the assignment of patients to treatments randomised?)

การสุ่ม (Randomisation) คือการเปิดโอกาสให้ผู้เข้าร่วมการศึกษาแต่ละคน มีโอกาสเท่าเทียมกันที่จะได้รับสิ่งแทรกแซงหรือสิ่งเปรียบเทียบ โดยนอกเหนือจากการสุ่มแล้ว การปกปิดผลการแบ่งกลุ่ม (allocation concealment) ก็มีส่วนสำคัญเป็นอย่างมาก ทั้งนี้เพื่อป้องกันอคติที่เกิดจากการคัดเลือกกลุ่มตัวอย่าง (selection bias) อย่างไรก็ตามมักมีความเข้าใจคลาดเคลื่อนว่าการปกปิดผลการแบ่งกลุ่มนั้นเป็นส่วนหนึ่งของการปกปิด (blinding หรือ masking)

แท้จริงแล้วกระบวนการทั้งสองมีความแตกต่างกัน โดยสิ้นเชิง กล่าวคือ การปกปิดผลการแบ่งกลุ่ม เป็นการป้องกันไม่ให้ผู้ที่มอบสิ่งแทรกแซงหรือสิ่งเปรียบเทียบให้แก่ผู้เข้าร่วมการศึกษา ทราบลำดับที่เกิดจากการสุ่ม เพื่อป้องกันอคติในการเลือกที่จะให้หรือไม่ให้สิ่งแทรกแซงกับผู้เข้าร่วมการศึกษาแต่ละราย ดังนั้นการปกปิดผลการแบ่งกลุ่มนี้จะสิ้นสุดทันทีที่ผู้เข้าร่วมการศึกษาได้รับสิ่งแทรกแซงเรียบร้อยแล้ว

ในขณะที่การปกปิด เป็นขั้นตอนที่จะเกิดขึ้นหลังจากที่ผู้เข้าร่วมการศึกษาได้รับสิ่งแทรกแซงหรือสิ่งเปรียบเทียบไปเรียบร้อยแล้ว ทั้งนี้เพื่อป้องกันไม่ให้เกิดอคติจากการตรวจ (detection bias) โดยแพทย์และอคติจากการปฏิบัติตน (performance bias) โดยผู้ป่วย ซึ่งการปกปิดนี้จะยังคงอยู่ไปตลอดจนการศึกษาเสร็จสิ้น

ตารางที่ 1 ตัวอย่างวิธีการสุ่มและการปกปิดผลจากการสุ่ม¹

กระบวนการ	ตัวอย่าง
1. การสุ่ม (Randomisation)	
1.1 Simple randomisation (unrestricted randomisation)	เป็นการสุ่มอย่างง่าย โดยไม่ได้จำกัดให้แต่ละกลุ่มมีจำนวนคนเท่ากัน วิธีพื้นฐาน เช่น การโยนเหรียญ การทอยลูกเต๋า การสับไพ่ วิธีที่ใช้โดยทั่วไป เช่น computer generated random number, การใช้ตาราง random numbers
1.2 Restricted randomisation	เป็นการสุ่มที่กำหนดให้คนในแต่ละกลุ่มมีจำนวนเท่ากัน เช่น - Block randomisation โดยขนาดของบล็อกอาจสุ่มเปลี่ยนไปเรื่อย ๆ เพื่อป้องกันไม่ให้นายสามารถทำนายลำดับที่เกิดจากการสุ่ม เรียกว่า random permuted blocks นอกจากนี้มีคำแนะนำว่าไม่ควรใช้ขนาดของบล็อกที่เล็กกว่า 6 เพื่อลดความเสี่ยงต่อการทำนายผลของการสุ่ม ⁹ - Stratified block randomisation เป็นการสุ่มที่แยกตามลักษณะที่สำคัญ เช่น หากแยกอันตรายภาคโดยเพศ (stratified by gender) จะแยกผู้เข้าร่วมการศึกษาที่เป็นชายและหญิงออกจากกัน แล้วทำการสุ่มเพื่อแบ่งกลุ่มแทรกแซงกับกลุ่มเปรียบเทียบแยกระหว่างชายกับหญิง จะทำให้มีสัดส่วนของเพศชายและหญิงที่ใกล้เคียงกันระหว่างกลุ่ม - Minimisation เป็นวิธีที่ถูกนำมาใช้กับการศึกษาที่มีขนาดเล็ก เพื่อลดความไม่สมดุลของลักษณะพื้นฐาน (baseline characteristics) ที่สำคัญระหว่างกลุ่ม โดยผู้เข้าร่วมการศึกษาคนแรก จะถูกสุ่มเข้ากลุ่มตามปกติ แต่ผู้เข้าร่วมการศึกษาคนต่อ ๆ มา จะถูกเลือกให้เข้ากลุ่มใดขึ้นกับลักษณะของผู้ป่วยคนนั้น ทั้งนี้ขอแนะนำให้ผู้อ่านที่สนใจศึกษารายละเอียดจากเอกสารอ้างอิง ¹⁰

ตารางที่ 1 (ต่อ)

กระบวนการ	ตัวอย่าง
2. การปกปิดผลการแบ่งกลุ่ม (Allocation concealment)	- Central randomisation เช่น interactive voice response systems (IVRS) หรือ interactive web response system (IWRS) - Independent central pharmacy - Telephone or internet-based randomisation service providers - Sequentially numbered drug containers - Sequentially numbered, opaque, sealed envelopes (SNOSE) ข้อพิจารณา: ผู้อ่านงานวิจัยควรพิจารณาว่านอกเหนือจากการปกปิดแล้ว ผู้วิจัยได้ระบุวิธีการปกปิดผลการแบ่งกลุ่มไว้ด้วยหรือไม่ หากไม่ได้ระบุไว้ อาจเพิ่มความเสี่ยงต่อการเกิดอคติจากการคัดเลือกกลุ่มตัวอย่าง นอกจากนี้ในปัจจุบันจะไม่แนะนำให้ใช้ SNOSE ในการปกปิดผลการแบ่งกลุ่ม เนื่องจากอาจเพิ่มความเสี่ยงต่อการคาดเดาผลการแบ่งกลุ่ม⁹ จนนำไปสู่อคติจากการคัดเลือกกลุ่มตัวอย่าง

3. ข้อสรุปของการศึกษาได้นับรวมผู้เข้าร่วม

การศึกษาทุกรายอย่างเหมาะสมหรือไม่ (Were all of the patients who entered the trial properly accounted for at its conclusion?)

ในข้อคำถามนี้มี 2 ประเด็นที่ต้องพิจารณา ได้แก่ ประเด็นที่หนึ่ง ในการศึกษาที่มีผู้ป่วยที่ออกจากการศึกษาก่อนกำหนดมาน้อยเพียงใด และมีสัดส่วนที่คล้ายคลึงกันระหว่างกลุ่มหรือไม่ เนื่องจากหากในการศึกษานั้นมีผู้ป่วยออกจากการศึกษาเป็นจำนวนมาก อาจทำให้ลักษณะของประชากรในกลุ่มศึกษาและกลุ่มเปรียบเทียบมีความแตกต่างกัน เนื่องจากคุณสมบัติจากการสุ่มหมดไป เกณฑ์ในการพิจารณาเบื้องต้นคือ จำนวนผู้ที่ออกจากการศึกษาก่อนกำหนดไม่ควรเกินร้อยละ 20 ของผู้เข้าร่วมการศึกษาทั้งหมด ทั้งนี้ หากเป็นผลลัพธ์ที่มีอุบัติการณ์ต่ำ ผู้ป่วยที่ออกจากการศึกษาแม้เพียงจำนวนน้อย ก็อาจทำให้ผลการศึกษามีอคติและไม่น่าเชื่อถือได้

ประเด็นที่สองที่ผู้อ่านต้องพิจารณา คือ ผู้ป่วยถูกวิเคราะห์ผลตามกลุ่มที่ถูกสุ่มมาตั้งแต่แรกหรือไม่ โดยสามารถพิจารณาจากว่าในการศึกษานั้นใช้การวิเคราะห์ผลแบบ Intention-to-treat (ITT) analysis หรือ Per-protocol analysis (PPA) โดยทั่วไปการเลือกวิธีการวิเคราะห์ผลลัพธ์ขึ้นอยู่กับรูปแบบการศึกษา กล่าวคือ การศึกษาชนิด equivalence trial จะเหมาะกับการวิเคราะห์แบบ PPA ในขณะที่การศึกษาชนิด superiority trial เหมาะกับการวิเคราะห์แบบ ITT สำหรับการศึกษาแบบ non-inferiority study นั้น มีคำแนะนำว่า ทั้ง ITT และ

PPA มีความสำคัญเท่ากัน จึงควรรายงานผลการวิเคราะห์จากทั้งสองวิธี¹¹⁻¹² นอกจากนี้ในบางการศึกษาอาจใช้วิธีวิเคราะห์ผลแบบ modified Intention-to-treat (mITT) analysis ซึ่งจะวิเคราะห์ผลเฉพาะผู้เข้าร่วมการศึกษาที่ได้รับสิ่งแทรกแซงมาแล้ว 1 ครั้งขึ้นไปเท่านั้น ก็มีหลักการพิจารณาลำยากับ ITT กล่าวคือ เหมาะกับการศึกษารูปแบบ superiority trial และควรรายงานร่วมกับ PPA สำหรับการศึกษาแบบ non-inferiority trial

4. มีการปกปิดผู้ป่วย บุคลากรทางการแพทย์ผู้ดูแลผู้ป่วย หรือผู้ทำการศึกษาหรือไม่ (Were patients, health workers and study personnel blind to treatment?)

ข้อคำถามนี้ผู้อ่านต้องพิจารณาว่าการศึกษาที่ผู้อ่านอ่าน มีการปกปิดลักษณะใดระหว่าง การปกปิดทางเดียว (single blind คือ การปกปิดผู้ป่วย) การปกปิดสองทาง (double blind คือ การปกปิดผู้ป่วยและแพทย์ผู้ให้สิ่งแทรกแซง) การปกปิดสามทาง (triple blind คือ การปกปิดผู้ป่วย แพทย์ผู้ให้สิ่งแทรกแซง และผู้ประเมินผลลัพธ์) หรือปกปิดเฉพาะผู้ประเมินผลลัพธ์ ในการศึกษาที่ไม่มีการปกปิดผู้ป่วยและแพทย์ผู้ดูแลผู้ป่วย หรือไม่สามารถปกปิดบุคคลดังกล่าวได้ เช่น การศึกษาเปรียบเทียบการทำหัตถการรูปแบบต่าง ๆ ควรวัดผลลัพธ์ที่เป็นปรนัย (objective outcome) และควรมีการปกปิดผู้ประเมินผลลัพธ์ร่วมด้วย

5. แต่ละกลุ่มที่เปรียบเทียบมีลักษณะที่คล้ายคลึงกัน ตั้งแต่เริ่มการศึกษาหรือไม่ (Were the groups similar at the start of the trial?)

ในข้อนี้ผู้อ่านควรพิจารณาว่า ในงานวิจัยนั้น ได้รายงานปัจจัยสำคัญที่สัมพันธ์กับการเกิดผลลัพธ์ไว้ครบถ้วนหรือไม่ ซึ่งมักได้จากการทบทวนข้อมูลเรื่องโรคและยาที่ได้กล่าวไปข้างต้น และหากมีการรายงานครบถ้วน ปัจจัยเหล่านั้นมีความคล้ายคลึงกันระหว่างกลุ่มหรือไม่ โดยให้พิจารณาจากลักษณะพื้นฐานที่มักแสดงในตารางที่ 1 ของบทความวิจัย และเปรียบเทียบความแตกต่างจากการพิจารณาองค์รวมโดยไม่ต้องใช้วิธีทางสถิติ เพื่อทดสอบความแตกต่างระหว่างกลุ่ม เนื่องจากการไม่พบนัยสำคัญทางสถิติ อาจเป็นผลมาจากอำนาจทางสถิติไม่เพียงพอ เพราะในการคำนวณขนาดตัวอย่างของการศึกษา เป็นการคำนวณเพื่อทดสอบความแตกต่างของการเกิดผลลัพธ์ ไม่ได้คำนวณเพื่อทดสอบความแตกต่างของลักษณะพื้นฐาน และในกรณีที่พบว่าลักษณะพื้นฐานมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ก็ไม่สามารถระบุได้ว่าเป็นนัยสำคัญทางสถิติที่เกิดจากความบังเอิญ หรือเป็นความแตกต่างที่แท้จริง ดังนั้นผลการเปรียบเทียบทางสถิติของลักษณะพื้นฐานในการศึกษารูปแบบ RCT จึงไม่ได้ให้ประโยชน์อะไรเพิ่มเติม และยังคงอาจสร้างความเข้าใจผิดแก่ผู้อ่านงานวิจัยได้

อนึ่ง แม้ว่าการสุ่มจะมีวัตถุประสงค์เพื่อลดอคติที่เกิดจากการคัดเลือก และเพื่อกระจายตัวแปรกวน (confounder) ให้เท่ากันระหว่างกลุ่มทดลองและกลุ่มควบคุม แต่ก็อาจมีโอกาสดังที่ลักษณะหรือปัจจัยบางอย่างจะมีความแตกต่างกันระหว่างกลุ่มได้ โดยเฉพาะอย่างยิ่งในการศึกษาขนาดเล็ก ประเด็นสำคัญที่ผู้อ่านควรพิจารณาต่อคือ หากมีลักษณะที่แตกต่างกันระหว่างกลุ่ม ผู้วิจัยมีวิธีการจัดการหรือควบคุมลักษณะดังกล่าวอย่างไร เช่น มีการปรับด้วย

วิธีทางสถิติ หรือมีการวิเคราะห์แบบแยกอันตรภาคชั้นตามลักษณะดังกล่าวหรือไม่ เป็นต้น

6. นอกเหนือจากสิ่งแทรกแซงในการทดลองแล้ว ผู้ป่วยแต่ละกลุ่มได้รับการรักษาร่วมที่คล้ายคลึงกันหรือไม่ (Aside from the experimental intervention, were the groups treated equally?)

ในข้อคำถามนี้ ผู้อ่านควรพิจารณาในส่วนของการเปรียบวิธีวิจัยหรือโปรโตคอลของงานวิจัยนั้น ว่ามีการควบคุมสิ่งแทรกแซงร่วมหรือการรักษาอื่นที่ผู้ป่วยแต่ละกลุ่มได้รับหรือไม่ อย่างไร โดยเฉพาะอย่างยิ่งในการศึกษาที่ไม่ได้มีการปกปิดแบบสองทาง อาจมีการให้สิ่งแทรกแซงร่วมที่แตกต่างกันระหว่างกลุ่มทดลองและกลุ่มควบคุม ที่หากไม่มีการควบคุม จะทำให้ผลการศึกษาไม่น่าเชื่อถือ เพราะจะไม่สามารถระบุได้ว่าผลลัพธ์ที่เกิดขึ้น เป็นผลมาจากสิ่งแทรกแซงที่ศึกษา หรือเป็นผลจากสิ่งแทรกแซงร่วมที่ผู้ป่วยได้รับ หรืออาจเป็นผลเสริมกันของทั้งสิ่งแทรกแซงและสิ่งแทรกแซงร่วม

ส่วนที่สอง: ผลการศึกษาคืออะไร (What are the results?)

7. ขนาดของผลลัพธ์เป็นอย่างไร (How large was the treatment effect?)

ในข้อนี้ผู้อ่านต้องพิจารณาว่าขนาดของผลลัพธ์ (effect size หรือ treatment effect) ในการศึกษานั้นมีนัยสำคัญทางคลินิกหรือไม่ ขนาดของผลลัพธ์ที่นิยมรายงานโดยทั่วไป ได้แก่ mean difference (กรณีที่ผลลัพธ์เป็นตัวแปรต่อเนื่อง) Odds Ratio (OR) และ Risk Ratio (หรือ Relative Risk; RR) (กรณีที่ผลลัพธ์เป็นตัวแปรชนิดแบ่งกลุ่มสอง) และ hazard ratio

(HR ในกรณีที่ผลลัพธ์เป็นชนิดแบ่งกลุ่มสองแต่มีการนำระยะเวลาติดตามผู้ป่วยมารวมวิเคราะห์ด้วย) โดยสิ่งที่พึงระลึกไว้เสมอคือ “**นัยสำคัญทางสถิติ (statistically significant) ไม่เท่ากับนัยสำคัญทางคลินิก (clinically significant)**” เนื่องจากในทางสถิติ การเพิ่มขนาดตัวอย่างให้ใหญ่ขึ้น จะช่วยให้สามารถตรวจพบความแตกต่างเพียงเล็กน้อยที่ไม่มีนัยสำคัญทางคลินิกได้ เรียกว่าเป็นหลุมพรางจากการมีกลุ่มตัวอย่างขนาดใหญ่ เช่น การศึกษาผลของยาลดความดันโลหิตชนิดใหม่เทียบกับยามาตรฐาน หากทดสอบในคนจำนวน 10,000 คนแล้วพบว่ายาใหม่มีประสิทธิภาพในการลดความดันโลหิตค่าบนลง 0.5 มิลลิเมตรปรอท (95%CI 0.1 to 0.9, $P < 0.001$) ในทางปฏิบัติย่อมไม่มีแพทย์ท่านใดสั่งใช้ยาดังกล่าว เพราะความดันโลหิตที่ลดลงเพียง 0.5 มิลลิเมตรปรอทไม่มีนัยสำคัญในทางคลินิกเป็นต้น

ดังนั้นการพิจารณาขนาดของผลลัพธ์ในแง่ของนัยสำคัญทางคลินิก จึงขึ้นอยู่กับโรคและบริบทอื่น ๆ แต่มีหลักในการพิจารณาเบื้องต้นดังนี้ กรณีผลลัพธ์เป็นตัวแปรต่อเนื่อง การเปลี่ยนแปลงของค่าอย่างน้อย ร้อยละ 20-30 จากค่าเริ่มต้นอาจพิจารณาว่านัยสำคัญทางคลินิก กรณีผลลัพธ์เป็นตัวแปรแบบแบ่งสอง ค่า OR และ RR ที่น้อยกว่า 0.5 หรือมากกว่า 2.0 อาจพิจารณาว่านัยสำคัญทางคลินิก อย่างไรก็ตามในกรณีที่เป็นการศึกษาโรคที่ร้ายแรง โดยที่สิ่งแทรกแซงนั้นไม่มีอันตรายรุนแรงและราคาไม่แพงมากนัก ค่า OR หรือ RR ที่เข้าใกล้ 1 ก็ยังอาจพิจารณาว่านัยสำคัญทางคลินิกได้เช่นกัน¹³ สำหรับ HR มีผู้เสนอว่าค่าที่มากกว่า 1.57 หรือน้อยกว่า 0.6 อาจใช้พิจารณาว่านัยสำคัญทางคลินิก¹⁴

8. ความแม่นยำของผลลัพธ์เป็นอย่างไร (How precise was the estimate of the treatment effect?)

ข้อนี้ให้ผู้อ่านพิจารณาความแม่นยำของผลลัพธ์ที่รายงาน โดยพิจารณาจากช่วงความเชื่อมั่น 95% (95% Confidence interval; CI) ว่ามีค่าไปในทิศทางเดียวกันหรือไม่ เช่น กรณี 95%CI ของความแตกต่างของค่าเฉลี่ยที่ครอบคลุมทั้งค่าลบและค่าบวก หรือ 95%CI ของสัดส่วน เช่น OR RR หรือ HR ที่ครอบคลุมค่าที่มากกว่าและน้อยกว่าหนึ่ง แสดงให้เห็นถึงความไม่แม่นยำของผลลัพธ์การศึกษา ทำให้ยากแก่การตัดสินใจว่า ผู้ป่วยจะเกิดประโยชน์จากสิ่งแทรกแซงที่ให้หรือไม่ นอกจากนี้ควรพิจารณาขอบเขตบนและขอบเขตล่างของช่วงความเชื่อมั่นว่ายังสะท้อนถึงนัยสำคัญทางคลินิกอยู่หรือไม่

ส่วนที่สาม: จะประยุกต์ใช้ผลการศึกษานี้กับผู้ป่วยได้อย่างไร (How can the results be applied to the patients?) หรือผลการศึกษาจะช่วยท้องถิ่นหรือไม่ (Will the results help locally?)

9. จะสามารถประยุกต์ใช้ผลลัพธ์จากการศึกษากับผู้ป่วยในท้องถิ่นหรือในบริบทของเราได้หรือไม่ (Can the results be applied to the local population, or in your context?)

ในข้อนี้ควรเริ่มพิจารณาว่าผู้ป่วยที่เราจะนำผลการศึกษาไปใช้นั้นมีลักษณะตรงกับเกณฑ์คัดเข้าในการศึกษานั้นหรือไม่ รวมทั้งไม่ได้มีลักษณะตามเกณฑ์ในการคัดออก หลังจากนั้นจึงพิจารณาลักษณะพื้นฐานของประชากรในการศึกษาว่ามีความคล้ายคลึงกับผู้ป่วยของเราหรือไม่ นอกจากนี้ผู้อ่านควรพิจารณาให้ถ่วงถ่วงก่อนนำผลของการวิเคราะห์กลุ่มย่อยไปใช้กับผู้ป่วย เนื่องจากการวิเคราะห์กลุ่มย่อยอาจทำให้พบนัยสำคัญทางสถิติได้โดยบังเอิญ

ซึ่งเป็นผลมาจากการเปรียบเทียบซ้ำหลายครั้ง (multiple comparisons) นอกจากนี้การแบ่งกลุ่มย่อย อาจทำให้ผลของการสุ่มหายไป โดยเฉพาะอย่างยิ่งหากการศึกษาไม่ได้มีการวางแผนที่จะสุ่มแยกตามอันตรายขั้นไว้ตั้งแต่ต้น กล่าวได้อีกนัยหนึ่งว่า หากเป็นการแบ่งกลุ่มย่อยที่มีการวางแผนไว้ตั้งแต่เริ่มการศึกษา และผลลัพธ์ที่ได้ถูกปรับการเปรียบเทียบซ้ำแล้วก็สามารถเชื่อถือได้

10. มีการพิจารณาผลลัพธ์ที่มีความสำคัญทางคลินิกไว้อย่างครบถ้วนหรือไม่ (Were all clinically important outcomes considered?)

ประเด็นของความครอบคลุมผลลัพธ์ทางคลินิกที่สำคัญ ผู้อ่านจะทราบได้จากการทบทวนข้อมูลเรื่องโรคและยาที่เกี่ยวข้อง

11. ประโยชน์ที่ได้รับมีความคุ้มค่ากับอันตรายและค่าใช้จ่ายที่อาจเพิ่มขึ้นหรือไม่ (Are the benefits worth the harms and costs?)

ในข้อนี้ผู้อ่านจะต้องคำนวณและเปรียบเทียบ number needed to treat (NNT) และ number needed to harm (NNH) จากสิ่งแทรกแซงที่ศึกษา สำหรับการคำนวณ NNT (หรือ NNH) ในกรณีที่มีการศึกษานั้นรายงานจำนวนผู้เกิดผลลัพธ์และจำนวนคนทั้งหมดในแต่ละกลุ่ม NNT (หรือ NNH) จะสามารถคำนวณได้จาก “หนึ่งส่วน absolute risk reduction” ดังจะได้อธิบายต่อไป

Risk หรือความเสี่ยง คือ จำนวนคนที่เกิดผลลัพธ์ต่อจำนวนคนทั้งหมดในแต่ละกลุ่ม ตัวอย่างเช่น ผลการศึกษา AUGUSTUS¹⁵ ซึ่งเป็นงานวิจัยรูปแบบ RCT โดยเปรียบเทียบ Apixaban (สิ่งแทรกแซง) กับ Vitamin K antagonist (VKA; สิ่งเปรียบเทียบ) ในผู้ป่วยโรคหัวใจห้องบนสั่นพริ้ว (atrial fibrillation; AF) ที่มีภาวะ ACS (acute coronary syndrome)

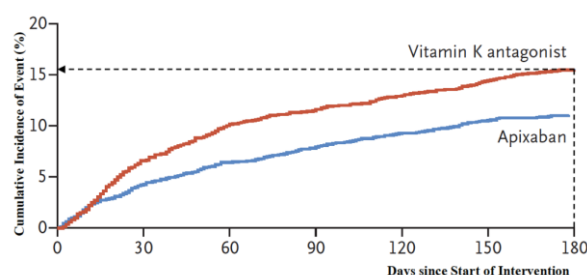
ร่วมด้วย โดยเปรียบเทียบการเกิดภาวะเลือดออก ณ เวลา 180 วัน พบว่าในกลุ่มที่ได้รับ apixaban มีคนที่เกิดภาวะเลือดออก 241 คน จากทั้งหมด 2,290 คน ดังนั้นความเสี่ยงของการเกิดภาวะเลือดออกในกลุ่มที่ได้รับ apixaban จะเท่ากับ 241/ 2,290 หรือ 0.105 ในขณะที่ความเสี่ยงของการเกิดภาวะเลือดออกในกลุ่มที่ได้รับ VKA จะมีผู้เกิดภาวะเลือดออก 332 คน จาก 2,259 คน สามารถคำนวณความเสี่ยงของการเกิดภาวะเลือดออกในกลุ่ม VKA ได้เท่ากับ 332/ 2,259 หรือ 0.147 เมื่อคำนวณความเสี่ยงของแต่ละกลุ่มได้แล้ว ให้นำความเสี่ยงของทั้งสองกลุ่มมาลบกันและหาค่าสัมบูรณ์ (absolute) จะได้ 0.105 – 0.147 หรือ -0.042 เท่ากับ 0.042 เมื่อคำนวณ NNT จะได้เป็น 1/0.042 หรือประมาณ 24

ในกรณีที่การศึกษานั้นไม่ได้รายงานจำนวนผู้เกิดผลลัพธ์ในแต่ละกลุ่ม จะสามารถคำนวณ NNT (หรือ NNH) ได้จาก hazard ratio และ survival probability ที่ประมาณจากกราฟ Kaplan-Meier โดยใช้สูตร¹⁶

$$NNT \text{ (or NNH)} = \frac{1}{(S_{tc}^{HR} - S_{tc})}$$

เมื่อ HR คือ Hazard ratio ของผลลัพธ์จากการศึกษา S_{tc} คือ Survival probability ของกลุ่มควบคุมที่เวลาใด ๆ

ตัวอย่าง Kaplan-Meier plot ของการเกิดภาวะเลือดออก ในศึกษา AUGUSTUS (รูปที่ 1)



รูปที่ 1 Kaplan-Meier plot ของการเกิดภาวะเลือดออกในการศึกษา AUGUSTUS

(ดัดแปลงจาก: Lopes RD, et al. N Engl J Med. 2019;380(16):1509-24.)¹⁵

ในบทความระบุว่าค่า hazard ratio ของการเกิดภาวะเลือดออก คือ 0.69 (95%CI 0.58-0.81) โดยจากรูปที่ 1 เมื่อลากเส้นตั้งฉากจากแกน X ที่ 180 วัน ขึ้นไปตัดเส้นกราฟของ VKA จากนั้นจึงลากเส้นตัดแกน Y จะได้ค่าประมาณ 15% นั่นคือ survival probability ที่ 180 วันของกลุ่มควบคุมเท่ากับ $100 - 15 = 85\%$ หรือ 0.85 (ในกรณีที่แกน Y เป็น survival probability หรือ incidence of event-free อยู่แล้วสามารถนำค่าดังกล่าวมาใช้ได้โดยไม่ต้องนำไปลบจาก 100%) ดังนั้นจะสามารถคำนวณ NNT ได้เท่ากับ $1/[(0.85^{0.69}) - (0.85)] = 22.7 \sim 23$ นั่นเอง ซึ่งแปลความได้ว่า จะต้องให้ apixaban แก่ผู้ป่วยประมาณ 23 คน เป็นระยะเวลา 180 วัน เพื่อป้องกันการเกิดภาวะเลือดออกในผู้ป่วย 1 คน วิธีการคำนวณ NNT จากกราฟ Kaplan-Meier นี้มีข้อดีคือ ผู้อ่านสามารถคำนวณ NNT ณ จุดเวลาต่าง ๆ ได้ตามที่ปรากฏในกราฟ นอกจากนี้ยังสามารถคำนวณช่วงความเชื่อมั่นของ NNT ได้ เช่น 95% CI ของตัวอย่างนี้คือ $1/[(0.85^{0.58}) - (0.85)]$ ถึง $1/[(0.85^{0.81}) - (0.85)]$ ซึ่งมีค่าเท่ากับ 17 ถึง 38 นั่นเอง นั่นคือ NNT เท่ากับ 23 (95%CI 17 to 38) ทั้งนี้ผู้อ่านพึงระวังความคลาดเคลื่อนที่เกิดจากประมาณ survival probability บนแกน Y ที่ไม่ได้มีมาตราส่วนระบุไว้อย่างชัดเจน

4. สรุปผลจากการประเมินงานวิจัย

ความตรงของผลการศึกษา: การสรุปความตรงของผลการศึกษาที่มีความเป็นอัตวิสัย (subjective) สูงขึ้นกับประสบการณ์และความเชี่ยวชาญของผู้ประเมิน จึงมักไม่ต้องการคำตอบว่า “มี” หรือ “ไม่มี” แต่จะเป็นการพิจารณาว่า ในภาพรวมแล้วระเบียบวิธีวิจัยที่ผู้วิจัยใช้มีความน่าเชื่อถือมากเพียงพอที่จะเชื่อผลลัพธ์ที่ผู้วิจัยสรุปมาหรือไม่

ผลการศึกษา: เป็นการสรุปผลลัพธ์ที่รายงานพร้อมช่วงความเชื่อมั่น โดยควรระบุว่าผลลัพธ์ที่ได้มีความแม่นยำและมีนัยสำคัญทางคลินิกหรือไม่ การประยุกต์ใช้ผลการศึกษากับผู้ป่วยหรือท้องถิ่น: เมื่อพิจารณาจากเกณฑ์การคัดเข้า การคัดออก และลักษณะของผู้เข้าร่วมการศึกษาแล้ว ผลการศึกษาดังกล่าว สามารถนำมาประยุกต์ใช้กับผู้ป่วยที่เราดูแลภายใต้บริบทของท้องถิ่นที่เราปฏิบัติงานหรือไม่ และสิ่งแทรกแซงที่สนใจมี NNT (หรือ NNH) ตลอดจนก่อให้เกิดผลไม่พึงประสงค์กับตัวผู้ป่วยอย่างไรบ้าง ในกรณีที่เป็นยาใหม่ ควรพิจารณาราคายาต่อระยะเวลาการรักษาด้วย

บทสรุป

แคลสปีเป็นเครื่องมือที่ใช้เพื่อประเมินความน่าเชื่อถือของงานวิจัยรูปแบบ RCT ซึ่งประกอบด้วยชุดคำถาม 11 ข้อ โดย 6 ข้อคำถามแรกจะช่วยให้การประเมินความตรงภายในของงานวิจัย 2 ข้อคำถามต่อมาจะเน้นการประเมินผลลัพธ์ของงานวิจัย และ 3 ข้อสุดท้ายจะช่วยให้การพิจารณาว่าจะสามารถประยุกต์ใช้ผลงานวิจัยกับผู้ป่วยได้อย่างไร ทั้งนี้แคลสปีเป็นเพียงหนึ่งในหลาย ๆ เครื่องมือที่ผู้อ่านงานวิจัยสามารถเลือกใช้ได้ สำหรับผู้ที่ยังมีประสบการณ์อ่านงานวิจัยรูปแบบ RCT ไม่มากนัก การฝึกใช้เครื่องมือแคลสปีจะช่วยจัดกระบวนการคิดให้เป็นระบบ ส่งผลให้สามารถบริโภคข้อมูลจากงานวิจัยได้อย่างมีประสิทธิภาพ

การประเมินความน่าเชื่อถือของงานวิจัยรูปแบบ RCT เป็นทักษะพื้นฐานสำคัญที่บุคลากรทางการแพทย์จะต้องฝึกฝนจนชำนาญ โดยเริ่มต้นจากการสรุปบทความให้เป็นเนื้อหาที่เข้าใจได้ง่าย มองเห็นภาพได้ชัดเจน ต่อมาจึงทบทวนข้อมูลของโรคและยาที่เกี่ยวข้อง

เพื่อกำหนดกรอบความคิดว่าการศึกษาคืออะไร ควรเป็นอย่างไร จากนั้นจึงประเมินงานวิจัยด้วยเครื่องมือใด ๆ ที่สุดท้ายผู้อ่านจะต้องสรุปให้ได้ว่าผลการศึกษานี้มีความถูกต้องน่าเชื่อถือหรือไม่ ผลการศึกษาเป็นอย่างไร และจะประยุกต์ใช้กับผู้ป่วยที่เราดูแลอย่างไร ในการฝึกประเมินงานวิจัยช่วงแรกอาจใช้เวลาพิจารณานาน แต่หากมีการฝึกฝนอย่างต่อเนื่อง ผู้อ่านจะสามารถอ่านและประเมินความน่าเชื่อถือได้อย่างรวดเร็ว ทั้งนี้เพื่อเป้าหมายในการดูแลผู้ป่วยอย่างมีประสิทธิภาพบนพื้นฐานของหลักฐานเชิงประจักษ์ทางการแพทย์นั่นเอง

กิตติกรรมประกาศ

ขอขอบพระคุณ อ. ดร. ภญ.สุธิดา บุญสม, อ. ดร. ภก.ยุทธนา วงศาสาณ และ อ. ภก.ศิริยุทธ พัฒนโสภณ ที่ช่วยตรวจทานและให้ข้อคิดเห็นที่เป็นประโยชน์ ทำให้บทความฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น

REFERENCES

1. 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.
2. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomised clinical trials: Is blinding necessary? Control Clin Trials. 1996;17(1):1–12.
3. Altman DG, Bland JM. Treatment allocation by minimisation. BMJ. 2005;330(7495):843.
4. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Reporting of noninferiority and equivalence randomised trials: extension of the CONSORT 2010 statement. JAMA. 2012;308(24):2594–604.
5. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA. 1993;270(21):2598–601.
6. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA. 1994;271(1):59–63.
7. Neal B, Perkovic V, Mahaffey KW, De Zeeuw D, Fulcher G, Erondou N, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med. 2017;377(7):644–57.
9. Powers AC. Chapter 419: Diabetes Mellitus - Complications. In: Kasper DL, Fauci AS, Hauser SL, Longo DL, Jameson JL, Loscalzo J, editors. Harrison's principles of internal medicine. 19th ed. New York: McGraw Hill Education Medical; 2015. p. 2422–30.

9. Kroon LA, Williams C. Diabetes mellitus. In: Alldredge BK, Jacobson PA, Corelli RL, Kradjan WA, Ernst ME, Williams BR, et al., editors. Applied therapeutics : The clinical use of drugs. 10th ed. Philadelphia:Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013. p. 1223–300.
10. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: Are we getting better? BMJ. 2016;355:1–5.
11. Facey KM, Lewis JA. Guidelines On Statistical Methods in Clinical Trials. In: Wiley StatsRef: Statistics Reference Online [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2014 [cited 2020 Mar 5]. Available from: <http://doi.wiley.com/10.1002/9781118445112.stat04940>
12. Points to consider on switching between superiority and non-inferiority. Br J Clin Pharmacol. 2001;52(3):223–8.
13. Streiner DL, Norman GR. Mine is bigger than yours measures of effect size in research. Chest. 2012;141(3):595–8.
14. Azuero A. A note on the magnitude of hazard ratios. Cancer. 2016;122(8):1298–9.
15. Lopes RD, Heizer G, Aronson R, Vora AN, Massaro T, Mehran R, et al. Antithrombotic therapy after acute coronary syndrome or PCI in atrial fibrillation. N Engl J Med. 2019;380(16):1509–24.
16. De Lemos ML, Barratt AL, Wyer PC, Guyatt G, Simpson JM. NNT for studies with long-term follow-up. CMAJ. 2005;172(5):613–5.

ตัวอย่างการประเมินงานวิจัยด้วยเครื่องมือแพทย์

1) สรุปงานวิจัย (รูปแบบขึ้นกับผู้อ่านเป็นสำคัญ)

2) ทบทวนข้อมูลเรื่องโรคและยาที่เกี่ยวข้องกับการศึกษา CANVAS program trial

โรคหลักที่สนใจในการศึกษานี้ คือโรคเบาหวานชนิดที่สองที่มีระยะเวลาการเป็นโรคยาวนานเกินกว่าสิบปี โดยผู้ป่วยส่วนใหญ่มีโรคหัวใจและหลอดเลือดร่วมด้วย และประมาณ 20-30% ของผู้ป่วยทั้งหมดจะมีภาวะแทรกซ้อนต่อหลอดเลือดขนาดเล็ก นอกจากนี้ผู้ป่วยส่วนใหญ่จะได้รับยา metformin และอินซูลินร่วมด้วยอยู่แล้วแต่ยังควบคุมระดับน้ำตาลในเลือดไม่ได้ ดังนั้นแนวทางการทบทวนความรู้เรื่องโรค เพื่อประเมินบทความนี้ จึงมุ่งเน้นการป้องกันทุติยภูมิ ซึ่งผลจากการสรุปงานวิจัยในขั้นตอนแรก จะทำให้ผู้อ่านเห็นว่าการศึกษานี้มุ่งเน้นการป้องกันโรคหัวใจและหลอดเลือดและชะลอการเสื่อมของไตเป็นหลัก

จากการทบทวนข้อมูลเรื่องโรคสามารถสรุปได้ว่า

- ภาวะไตเสื่อมและการเกิดภาวะแทรกซ้อนต่อระบบหัวใจและหลอดเลือดในผู้ป่วยเบาหวานมักพบหลังจากเริ่มป่วยด้วยโรคเบาหวานมาแล้วอย่างน้อย 10 ปี และจะใช้ระยะเวลาประมาณ 5 ปีกว่าที่ microalbuminuria จะพัฒนาไปเป็น macroalbuminuria⁷⁻⁸
- การชะลอการเสื่อมของไตจากโรคเบาหวาน ประกอบด้วย การควบคุมระดับน้ำตาลในเลือด การควบคุมความดันโลหิต และการให้ยาในกลุ่ม Angiotensin converting enzyme inhibitors (ACEIs) หรือ Angiotensin receptor blockers (ARBs)⁷
- ปัจจัยที่สัมพันธ์กับการเสียชีวิตและทุพพลภาพในผู้ป่วยเบาหวาน ได้แก่ เพศ โรคไขมันในเลือดผิดปกติ โรคความดันโลหิตสูง โรคอ้วน การมีกิจกรรมทางกายที่ลดลง การสูบบุหรี่ ภาวะ albuminuria การเพิ่มระดับ serum creatinine และการทำงานของเกล็ดเลือดและผนังหลอดเลือดที่ผิดปกติ⁷
- ในผู้ป่วยเบาหวาน อาจใช้ระยะเวลายาวนานกว่า 10 ปี กว่าที่การควบคุมระดับน้ำตาลในเลือดจะส่งผลดีต่อระบบหัวใจและหลอดเลือด⁹

จากข้อมูลที่สรุปข้างต้น ช่วยกำหนดกรอบแนวคิดก่อนการประเมินงานวิจัยได้ว่า

- ควรกำหนดระยะเวลาติดตามอย่างน้อย 5 ปี สำหรับติดตามการชะลอการดำเนินไปของโรคไต และอย่างน้อย 10 ปี ในการติดตามอัตราประโยชน์ต่อระบบหัวใจและหลอดเลือด
- ปัจจัยสำคัญที่ควรคล้ายคลึงกันระหว่างกลุ่ม เนื่องจากมีความสัมพันธ์กับการตายด้วยโรคหัวใจและหลอดเลือดในผู้ป่วยเบาหวาน ได้แก่ สัดส่วนของเพศชายต่อหญิง ระดับน้ำตาลในเลือด ระดับความดันโลหิต การได้รับยาในกลุ่ม ACEIs หรือ ARBs ระดับความดันโลหิต ดัชนีมวลกาย การสูบบุหรี่ และภาวะ albuminuria

3) ประเมินงานวิจัยที่ใช้เครื่องมือแพทย์

ข้อคำถามที่ประเมิน	การประเมิน	ข้อเสนอแนะ
Are the results valid?		
1. Did the trial address a clearly focused issue?	Yes	มีความชัดเจน กล่าวคือเป็นการศึกษาผลของยา Canagliflozin ขนาด 100-300 มก.ต่อวัน เปรียบเทียบกับยาหลอก ในผู้ป่วยเบาหวานชนิดที่ 2 ที่มีประวัติเป็นโรคหัวใจและหลอดเลือดหรือมีความเสี่ยงสูงต่อการเกิดโรคหัวใจและหลอดเลือด เพื่อติดตามประสิทธิผลหลักต่อ การตายจากโรคหัวใจและหลอดเลือด การเกิดโรคกล้ามเนื้อหัวใจขาดเลือด และการเกิดโรคหลอดเลือดสมอง และประสิทธิผลรองต่อการดำเนินไปของโรคไต และการเข้าโรงพยาบาลด้วยภาวะหัวใจล้มเหลว
2. Was the assignment of patients to treatments randomised?	Yes	ใน methods (หน้า 645) ระบุว่า "Randomisation was performed centrally through an interactive Web-based response system with the use of a computer-generated randomisation schedule with randomly permuted blocks..."
3. Were all of the patients who entered the trial properly accounted for at its conclusion?	Yes	หน้า 23 (Supplementary appendix) ระบุว่าผู้ที่ออกไปจากการศึกษาก่อนกำหนด คิดเป็น $184 + 224 = 408$ หรือ 4% ของผู้เข้าร่วมการศึกษาทั้งหมด ซึ่งเป็นสัดส่วนที่ไม่สูงมากนัก (<20%) นอกจากนี้ใช้วิธีวิเคราะห์ผลแบบ ITT เมื่อพิจารณาพร้อมกับผลลัพธ์หลักที่เกิดสูงถึง ~ 20% (จาก Kaplan-Meier Figure 2A หน้า 651) การใช้ ITT analysis ไม่น่าจะให้ผลที่แตกต่างจาก PPA มากนัก แม้ว่าผู้วิจัยจะตั้งสมมติฐานของ non-inferiority ไว้เป็นหลักก็ตาม
4. Were patients, health workers and study personnel blind to treatment?	Yes	หน้า 646 ส่วน methods ระบุว่า "Participants and all trial staff were unaware of the individual treatment assignments until completion of the trial." แสดงว่าอย่างน้อยต้องมีการปกปิดแบบสองทาง

ข้อคำถามที่ประเมิน	การประเมิน	ข้อเสนอแนะ
5. Were the groups similar at the start of the trial?	Yes	หน้า 648 Table 1 และ Table S4 ใน supplementary appendix เมื่อพิจารณาแล้วพบว่าลักษณะพื้นฐานระหว่างกลุ่มที่ได้รับสิ่งแทรกแซงและกลุ่มที่ได้รับยาหลอกมีความใกล้เคียงกัน รวมถึงปัจจัยที่มีผลต่อการตายด้วยโรคหัวใจและหลอดเลือดในผู้ป่วยเบาหวานชนิดที่สอง ได้แก่ เพศ HbA1c ความดันโลหิต ดัชนีมวลกาย การสูบบุหรี่ การได้รับยา ACEIs หรือ ARBs และการมีภาวะ albuminuria พบว่ามีความใกล้เคียงกันของทั้งสองกลุ่ม
6. Aside from the experimental intervention, were the groups treated equally?	No	หน้า 646 ส่วน methods ระบุว่า "Use of other background therapy for glycemic management and other control of risk factors were guided by best practice instituted in line with local guidelines." เป็นนัยว่าไม่ได้มีการควบคุมสิ่งแทรกแซงร่วม อย่างไรก็ตามการปกปิดแบบสองทางอาจช่วยลด อคติจากข้อมูล (information bias) ของทั้งแพทย์และผู้ป่วยได้ ประเด็นนี้จึงอาจไม่ต้องกังวลมากนัก
What are the results?		
7. How large was the treatment effect?	Not applicable*	เมื่อพิจารณาเฉพาะขนาดของผลลัพธ์สามารถประเมินในเบื้องต้นได้ว่า ขนาดของผลลัพธ์ทั้งหมดยังไม่ใหญ่มากนัก ($HR > 0.6$) แต่ผลลัพธ์ที่สนใจเป็นผลลัพธ์ที่รุนแรงและสำคัญในทางคลินิก จึงต้องพิจารณาเรื่องต้นทุนและความปลอดภัยของ canagliflozin ร่วมด้วยเพื่อตัดสินว่ามีความสำคัญทางคลินิกหรือไม่
8. How precise was the estimate of the treatment effect?	Not applicable*	95%CI ของผลลัพธ์หลักมีค่าระหว่าง 0.75 ถึง 0.97 แสดงว่าค่อนข้างแม่นยำ (ช่วงไม่กว้างนัก) และเป็นไปในทางเดียวกัน อย่างไรก็ตามขอบเขตบนและล่างก็อาจไม่มีความสำคัญทางคลินิก ทั้งนี้ 95%CI ของผลลัพธ์อื่นพบว่าไม่แม่นยำ (ยกเว้นโรคหัวใจล้มเหลวและ albuminuria) เพราะมีค่าที่กว้างและมีขอบเขตบนที่ > 1 สะท้อนว่าความแม่นยำของผลลัพธ์หลักที่เพิ่มขึ้นนี้ เป็นผลมาจากการเพิ่มอำนาจทางสถิติโดยการรวมผลลัพธ์ย่อยเข้าด้วยกัน ทำให้มีเหตุการณ์รวมมากขึ้น และมีช่วงความเชื่อมั่นที่แคบลงจนพบความสำคัญทางสถิติ

ข้อคำถามที่ประเมิน	การประเมิน	ข้อเสนอแนะ
How can the results be applied to the patients? (or will the results help locally?)		
9. Can the results be applied to the local population, or in your context?	Can't tell	การจะประเมินข้อนี้ได้ ขึ้นกับลักษณะของผู้ป่วยที่จะนำ การศึกษานี้ไปประยุกต์ โดยต้องเป็นผู้ป่วยเบาหวานชนิดที่ 2 มายาวนาน ที่มีประวัติเป็นโรคหัวใจและหลอดเลือดร่วมด้วย หรือมีความเสี่ยงสูงต่อการเกิดโรคหัวใจและหลอดเลือดโดยให้ พิจารณาจากเกณฑ์การคัดเข้าและคัดออก (Table S1 หน้า 27 ใน supplementary appendix) และลักษณะพื้นฐาน (Table1 หน้า 648)
10. Were all clinically important outcomes considered?	No	การศึกษานี้รายงานผลลัพธ์ซึ่งครอบคลุมภาวะแทรกซ้อนที่พบ ได้บ่อยในผู้ป่วยเบาหวานชนิดที่สอง แต่ไม่ได้มีการรายงาน ผลต่อภาวะแทรกซ้อนต่อหลอดเลือดขนาดเล็กอื่น ๆ เช่น โรคตา โรคระบบประสาท อย่างไรก็ตาม ผู้ป่วยเป็นเบาหวาน ชนิดที่ 2 มายาวนานกว่า 10 ปี และ 1/3 ของผู้เข้าร่วม การศึกษามีภาวะเหล่านี้มาตั้งแต่ต้น จึงอาจมีพยากรณ์ต่อ ภาวะแทรกซ้อนต่อหลอดเลือดขนาดเล็กที่ไม่ได้มาตั้งแต่แรก การไม่รายงานผลลัพธ์ดังกล่าว อาจไม่ส่งผลต่อการตัดสินใจ ให้ยากับผู้ป่วยมากนัก
11. Are the benefits worth the harms and costs?	Can't tell	Cumulative event in control at the end of study ~ 18% (Figure 2 A) -> Survival prob. in control = 82% (0.82), HR = 0.86 (0.75 to 0.97) คำนวณ NNT ได้เท่ากับ 42 (95%CI 24 to 204) หมายความว่า เมื่อรักษาผู้ป่วย 42 คน ด้วย canagliflozin แล้วติดตามเป็นระยะเวลา 2.4 ปี จะป้องกันการเกิดผลลัพธ์หลักได้ 1 คน แต่ทั้งนี้พบว่า canagliflozin เพิ่มความเสี่ยงต่อการถูกตัดเท้า การเกิด กระดูกหัก การเกิดภาวะติดเชื้อที่อวัยวะเพศ และความเสี่ยงต่อ ภาวะสูญเสียจากการขับปัสสาวะที่มากเกินไป อย่างไรก็ตาม canagliflozin อาจมีประโยชน์เพิ่มเติม ต่อการลดการเข้า โรงพยาบาล ด้วยภาวะหัวใจล้มเหลว และชะลอการดำเนินไป ของภาวะ albuminuria ได้ การเลือกใช้จึงควรพิจารณา โรคร่วมของผู้ป่วยประกอบ

*เป็นข้อคำถามที่ไม่ต้องตอบด้วย yes no หรือ can't tell เนื่องจากขึ้นต้นด้วย how

4) สรุปผลจากการประเมินงานวิจัย

ผลการศึกษา CANVAS Program โดยภาพรวมมีความน่าเชื่อถือ เพราะมีคำถามงานวิจัยที่ชัดเจน และมีระเบียบวิธีวิจัยที่ดี กล่าวคือ มีการสุ่มและการปกปิดผลของการแบ่งกลุ่มที่เหมาะสม มีการปกปิดแบบสองทาง ผู้เข้าร่วมการศึกษามีอัตราการออกจากการศึกษาที่กำหนดที่ไม่สูงมากนัก (4%) และใกล้เคียงกันระหว่างกลุ่ม นอกจากนี้ลักษณะที่มีความสัมพันธ์กับผลลัพธ์ก็มีการรายงานไว้ครบถ้วน และมีความใกล้เคียงกันระหว่างกลุ่ม อย่างไรก็ตามในการศึกษานี้ไม่ได้มีการควบคุมสิ่งแทรกแซงร่วม แต่การมีการปกปิดแบบสองทาง อาจช่วยลดอคติจากการตรวจวินิจฉัยของแพทย์และอคติจากการปฏิบัติตนของผู้ป่วยได้ ประเด็นนี้จึงอาจไม่ส่งผลกระทบต่อมากนัก ดังนั้นจึงสามารถพิจารณาขนาดและความแม่นยำของผลลัพธ์ รวมถึงการประยุกต์ใช้ผลลัพธ์จากการศึกษานี้ต่อไปได้

เมื่อพิจารณาผลลัพธ์ของการศึกษา พบว่าขนาดของผลลัพธ์หลักยังไม่ใหญ่มากนัก และความแม่นยำของผลลัพธ์หลักที่เพิ่มขึ้น ก็เป็นผลสืบเนื่องมาจากการรวมผลลัพธ์ย่อยเข้าด้วยกัน ทั้งนี้อาจเป็นไปได้ว่าระยะเวลาในการศึกษา

(ประมาณ 2.4 ปี) สั้นเกินกว่าที่จะพบอรรถประโยชน์ต่อหัวใจและหลอดเลือดจากยา อย่างไรก็ตามในแง่ของการชะลอภาวะ albuminuria และการเข้ารับการรักษาตัวในโรงพยาบาลด้วยภาวะหัวใจล้มเหลว แม้ขนาดของผลลัพธ์จะไม่ใหญ่นัก แต่ก็มีช่วงความเชื่อมั่นที่ค่อนข้างแม่นยำและเป็นไปในทิศทางเดียวกัน ทำให้ผลของ canagliflozin ต่อโรคหัวใจล้มเหลวและโรคไต เป็นผลลัพธ์ที่น่าจับตามองต่อไป

ในแง่ของการนำไปใช้ เบื้องต้นควรเป็นผู้ป่วยเบาหวานชนิดที่ 2 ที่ถูกวินิจฉัยมายาวนานและมีโรคหัวใจและหลอดเลือดร่วมด้วย ดังนั้น canagliflozin ไม่ควรถูกพิจารณาเป็นยาทางเลือกแรกในผู้ป่วยที่เพิ่งได้รับการวินิจฉัย นอกจากนี้ยังเพิ่มความเสี่ยงต่อการถูกตัดเท้า การเกิดกระดูกหัก และการติดเชื้อที่อวัยวะเพศ ดังนั้นควรติดตามผู้ป่วยในประเด็นจำเพาะ เช่น ติดตามแผลที่เท้าอย่างใกล้ชิด เป็นต้น สำหรับบทบาทของยาในการป้องกันแบบปฐมภูมิหรือทุติยภูมิต่อโรคหัวใจล้มเหลวยังไม่ชัดเจนนัก เนื่องจากมีผู้ป่วยร้อยละ 14 ที่ได้รับการวินิจฉัยด้วยโรคหัวใจล้มเหลวมาตั้งแต่เริ่มต้นการศึกษา

แบบทดสอบการศึกษาต่อเนื่อง สำหรับผู้ประกอบการวิชาชีพเภสัชกรรม

รหัสกิจกรรมการศึกษาต่อเนื่อง 1011-1-000-002-11-2563

จำนวน 3 หน่วยกิต

วันที่รับรอง : 9/11/2563

วันที่หมดอายุ : 8/11/2564

คำชี้แจง จงเลือกคำตอบที่ถูกต้องที่สุด (3 หน่วยกิตการศึกษาต่อเนื่อง)

1. ข้อใด ไม่ใช่ ปัจจัยที่ใช้พิจารณาคุณภาพของงานวิจัยรูปแบบ RCT
 - ก. Bias
 - ข. Confounder
 - ค. Impact factor
 - ง. Internal validity
 - จ. External validity
2. ข้อใด ไม่ใช่ ข้อคำถามสำคัญที่ใช้ประเมินคุณภาพงานวิจัยรูปแบบ RCT
 - ก. Are the results valid?
 - ข. What are the results?
 - ค. Will the results help locally?
 - ง. How can I apply the results to the patients?
 - จ. ทุกข้อคือคำถามที่สำคัญที่ใช้ประเมินคุณภาพงานวิจัยรูปแบบ RCT
3. ต่อไปนี้คือขั้นตอนการประเมินคุณภาพงานวิจัยรูปแบบ RCT ยกเว้น
 - ก. สืบค้นงานวิจัยในหัวข้อที่เกี่ยวข้องอย่างเป็นระบบ
 - ข. เขียนสรุปบทความตาม PICO
 - ค. ทบทวนข้อมูลเรื่องโรคและยาที่เกี่ยวข้อง
 - ง. ประเมินคุณภาพงานวิจัยตามเครื่องมือที่เลือกใช้
 - จ. สรุปคำตอบของข้อคำถามสำคัญในการประเมินคุณภาพงานวิจัย
4. ข้อใดกล่าวถูกต้องเกี่ยวกับความตรง (validity) ของการศึกษารูปแบบ RCT
 - ก. External validity เรียกอีกชื่อว่า generalisability
 - ข. การศึกษาที่มี external validity ไม่จำเป็นต้องมี internal validity
 - ค. การศึกษาที่มี internal validity ต้องมี external validity ร่วมด้วยเสมอ
 - ง. Internal validity เกิดจากความสามารถในการขยายผลการศึกษาไปสู่ประชากรทั่วไปได้
 - จ. External validity เกิดจากการออกแบบการศึกษาที่ดีและใช้วิธีวิเคราะห์ที่เหมาะสม ทำให้ผลการศึกษาปราศจากอคติและตัวแปรกวน

5. ข้อใดกล่าวถูกต้องเกี่ยวกับการสุ่ม (randomisation) และการปกปิดการแบ่งกลุ่ม (allocation concealment)
 - ก. การสุ่มจะทำให้กลุ่มที่ศึกษามีขนาดเท่ากันเสมอ
 - ข. การปกปิดการแบ่งกลุ่มเป็นส่วนหนึ่งของการปกปิด (blinding)
 - ค. ผลของการปกปิดการแบ่งกลุ่มจะยังคงอยู่ไปจนเสร็จสิ้นการศึกษา
 - ง. การสุ่มมีเป้าหมายเพื่อกระจายตัวแปรกวนให้เท่ากันระหว่างกลุ่มที่ศึกษา
 - จ. หากทดสอบทางสถิติแล้วพบว่ากลุ่มที่เปรียบเทียบกันมีลักษณะที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แสดงว่าสุ่มได้เหมาะสม
6. ตัวเลือกใดต่อไปนี้ ไม่ เข้าพวก
 - ก. Minimisation
 - ข. Central randomisation
 - ค. Unrestricted randomisation
 - ง. Stratified block randomisation
 - จ. Computer generated random number
7. ข้อใดกล่าวถูกต้องเกี่ยวกับประเด็นพิจารณาเรื่องจำนวนผู้เข้าร่วมการศึกษาตลอดช่วงการศึกษา
 - ก. การวิจัยในรูปแบบ superiority trial ควรวิเคราะห์ผลการศึกษาด้วยวิธี per protocol
 - ข. ผู้ที่ออกจากการศึกษาควรมีสัดส่วนคล้ายคลึงกันระหว่างกลุ่มทดลองและกลุ่มควบคุม
 - ค. หากผลลัพธ์มีอุบัติการณ์ต่ำ สัดส่วนผู้ที่ออกจากการศึกษาที่กำหนดไม่ควรเกินร้อยละ 20
 - ง. การวิเคราะห์ผลการศึกษาด้วยวิธี modified intention-to-treat อาจทำให้เกิดอคติ จึงเป็นวิธีที่ควรหลีกเลี่ยง
 - จ. การวิเคราะห์ผลโดยใช้ intention-to-treat จะช่วยขจัดอคติในกรณีที่ผู้เข้าร่วมการศึกษามีจำนวนมาก ออกจากการศึกษาก่อนกำหนด
8. ข้อใดกล่าวถูกต้องเกี่ยวกับการปกปิด (blinding)
 - ก. การศึกษาที่ดี จะต้องทำ double blinding เสมอ
 - ข. การปกปิดเป็นกระบวนการที่ช่วยป้องกัน selection bias
 - ค. หากไม่มีการปกปิด ควรวัดผลลัพธ์ที่เป็น subjective outcome
 - ง. หากไม่ได้ปกปิดแพทย์หรือผู้ป่วย ควรปกปิดผู้ประเมินผลลัพธ์
 - จ. กล่าวถูกทุกข้อ

9. ข้อใดกล่าวถูกต้องเกี่ยวกับการพิจารณาความคล้ายคลึงกันของลักษณะพื้นฐานระหว่างกลุ่มศึกษาและกลุ่มเปรียบเทียบ
- ควรใช้สถิติทดสอบร่วมด้วยเสมอ
 - หากพบลักษณะที่มีความแตกต่างกันระหว่างกลุ่ม แสดงว่าการสุ่มยังไม่เหมาะสม
 - ลักษณะระหว่างกลุ่มศึกษาและกลุ่มเปรียบเทียบไม่ควรแตกต่างกันอย่างมีนัยสำคัญทางสถิติ
 - หากในงานวิจัยมีการสุ่ม ผู้อ่านไม่จำเป็นต้องพิจารณาลักษณะพื้นฐานของกลุ่มศึกษาและกลุ่มเปรียบเทียบ
 - หากพบลักษณะที่แตกต่างกันระหว่างกลุ่ม ผู้วิจัยควรทดสอบความสัมพันธ์ของลักษณะดังกล่าวกับผลลัพธ์ และมีการปรับด้วยวิธีที่เหมาะสม
10. ในการศึกษา IDEAL (Incremental Decrease in End Points Through Aggressive Lipid Lowering) ซึ่งเปรียบเทียบการให้ Atorvastatin ขนาด 20 มก.ต่อวัน กับการให้ Simvastatin ขนาด 40 มก.ต่อวัน ในผู้ป่วยที่มีประวัติเป็นโรคกล้ามเนื้อหัวใจขาดเลือด โดยผู้วิจัยได้อธิบายในส่วนระเบียบวิธีวิจัยไว้ว่า “IDEAL is a multicenter, randomised, open-label study with stratified randomisation carried out in blocks by center and with blinded end point classification” (ที่มา: Pedersen TR, et al. JAMA. 2004;94(6):720-4.)
- หากพิจารณาเฉพาะข้อมูลข้างต้น ข้อใดกล่าวถูกต้อง
- ไม่มีวิธีลดอคติจากข้อมูล (information bias)
 - ไม่มีการปกปิดแพทย์ผู้ให้การรักษาและผู้ป่วย
 - มีการปกปิดผลการแบ่งกลุ่ม (allocation concealment)
 - จำนวนคนที่ได้รับ atorvastatin และ simvastatin อาจแตกต่างกันมาก
 - ถูกทุกข้อ

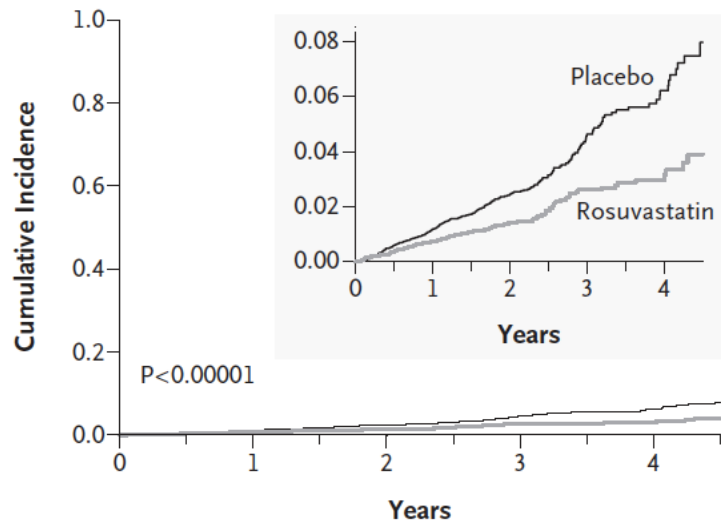
11. จากข้อมูลเป็นผลลัพธ์จากการศึกษา RCT เปรียบเทียบระหว่าง rosuvastatin กับยาหลอก ข้อใดกล่าวถูกต้อง

Table 3. Outcomes According to Study Group.						
End Point	Rosuvastatin (N=8901)		Placebo (N=8901)		Hazard Ratio (95% CI)	P Value
	No. of Patients	Rate per 100 person-yr	No. of Patients	Rate per 100 person-yr		
Primary end point	142	0.77	251	1.36	0.56 (0.46–0.69)	<0.00001
Nonfatal myocardial infarction	22	0.12	62	0.33	0.35 (0.22–0.58)	<0.00001
Any myocardial infarction	31	0.17	68	0.37	0.46 (0.30–0.70)	0.0002
Nonfatal stroke	30	0.16	58	0.31	0.52 (0.33–0.80)	0.003
Any stroke	33	0.18	64	0.34	0.52 (0.34–0.79)	0.002
Arterial revascularization	71	0.38	131	0.71	0.54 (0.41–0.72)	<0.0001
Hospitalization for unstable angina	16	0.09	27	0.14	0.59 (0.32–1.10)	0.09
Arterial revascularization or hospitalization for unstable angina	76	0.41	143	0.77	0.53 (0.40–0.70)	<0.00001
Myocardial infarction, stroke, or confirmed death from cardiovascular causes	83	0.45	157	0.85	0.53 (0.40–0.69)	<0.00001
Death from any cause						
Death on known date	190	0.96	235	1.19	0.81 (0.67–0.98)	0.03
Any death	198	1.00	247	1.25	0.80 (0.67–0.97)	0.02

(ที่มา: Ridker PM, et al. N Engl J Med. 2008;359(21):2195-207.)

- ก. ขนาดผลลัพธ์ต่อ primary end point มีนัยสำคัญทางคลินิกเนื่องจาก P-value มีค่า <0.00001
- ข. Hospitalization for unstable angina ไม่มีนัยสำคัญทางสถิติ แต่มีนัยสำคัญทางคลินิก
- ค. ความเสี่ยงของการเกิด primary end point ในกลุ่ม rosuvastatin เท่ากับ 0.77
- ง. ความเสี่ยงของการเกิด primary end point ในกลุ่ม placebo เท่ากับ 2.82%
- จ. กล่าวถูกทุกข้อ

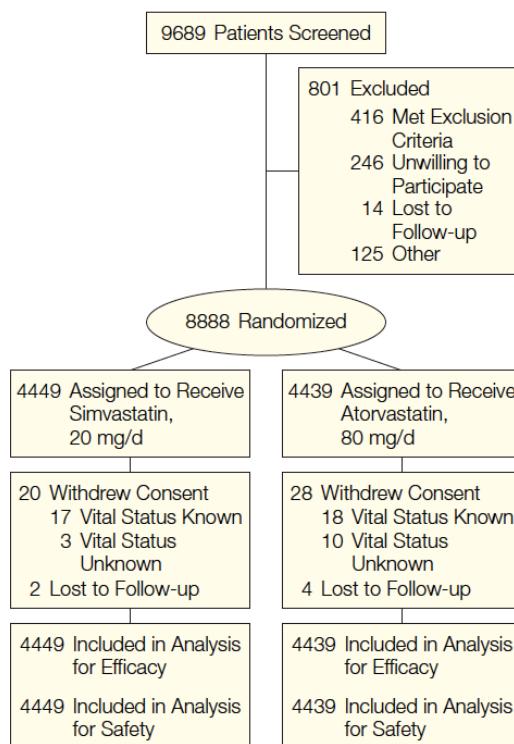
12. จากข้อมูลต่อไปนี้ จงคำนวณว่าต้องให้ rosuvastatin แก่ผู้ป่วยกี่คน จึงจะป้องกันการเกิด primary endpoint ณ เวลา 4 ปี ได้หนึ่งคน โดย hazard ratio ของการเกิด primary endpoint ในกลุ่ม rosuvastatin เปรียบเทียบกับกลุ่มที่ได้รับยาหลอกเท่ากับ 0.56 (95%CI 0.46 to 0.69)



(ที่มา: Ridker PM, et al. N Engl J Med. 2008;359(21):2195-207.)

- ก. 7
 - ข. 9
 - ค. 39
 - ง. 76
 - จ. ไม่มีข้อใดถูก
13. หากต้องการพิจารณาว่าลักษณะของผู้ป่วยที่จะนำผลการศึกษาไปใช้ ตรงกับลักษณะของผู้เข้าร่วมการศึกษาหรือไม่ ควรพิจารณาจากตัวเลือกต่อไปนี้ ยกเว้น
- ก. Subgroup analysis
 - ข. Exclusion criteria
 - ค. Inclusion criteria
 - ง. Baseline characteristics
 - จ. ควรพิจารณาจากทุกตัวเลือกข้างต้น

14. จาก participants flow diagram ของการศึกษา RCT เปรียบเทียบระหว่างการให้ Atorvastatin 80 mg/day (active group) และการให้ Simvastatin 20 mg/day (control group) ข้อใดกล่าวถูกต้อง



(ที่มา: Pedersen TG, et al. JAMA. 2005;294:2437-45.)

- ก. การศึกษานี้ใช้วิธีวิเคราะห์ผลลัพธ์แบบ per-protocol
- ข. การศึกษานี้มีผู้ออกจากการศึกษาก่อนกำหนดประมาณร้อยละ 8
- ค. การศึกษานี้มีสัดส่วนของผู้ออกจากการศึกษาก่อนกำหนดไม่สูงมากนัก
- ง. สัดส่วนของผู้ออกจากการศึกษาก่อนกำหนดระหว่างกลุ่มมีความแตกต่างกันสูง
- จ. ถูกทุกข้อ

15. ข้อใดกล่าวถูกต้อง

Outcome Measures	Simvastatin, No. (%) (n = 4449)	Atorvastatin, No. (%) (n = 4439)	Hazard Ratio (95% CI)	P Value
Major coronary event (primary outcome)	463 (10.4)	411 (9.3)	0.89 (0.78-1.01)	.07
CHD death	178 (4.0)	175 (3.9)	0.99 (0.80-1.22)	.90
Nonfatal myocardial infarction	321 (7.2)	267 (6.0)	0.83 (0.71-0.98)	.02
Cardiac arrest with resuscitation	7 (0.2)	10 (0.2)		
Any CHD event	1059 (23.8)	898 (20.2)	0.84 (0.76-0.91)	<.001
Coronary revascularization	743 (16.7)	579 (13.0)	0.77 (0.69-0.86)	<.001
Hospitalization for unstable angina	235 (5.3)	196 (4.4)	0.83 (0.69-1.01)	.06
Fatal or nonfatal stroke	174 (3.9)	151 (3.4)	0.87 (0.70-1.08)	.20
Major cardiovascular event*	608 (13.7)	533 (12.0)	0.87 (0.78-0.98)	.02
Hospitalization for nonfatal CHF	123 (2.8)	99 (2.2)	0.81 (0.62-1.05)	.11
Peripheral arterial disease†	167 (3.8)	127 (2.9)	0.76 (0.61-0.96)	.02
Any cardiovascular event	1370 (30.8)	1176 (26.5)	0.84 (0.78-0.91)	<.001
All-cause mortality	374 (8.4)	366 (8.2)	0.98 (0.85-1.13)	.81
Cardiovascular	218 (4.9)	223 (5.0)	1.03 (0.85-1.24)	.78
Noncardiovascular	156 (3.5)	143 (3.2)	0.92 (0.73-1.15)	.47
Malignant disease	112 (2.5)	99 (2.2)	0.89 (0.68-1.16)	.38
Suicide/violence/accidental death	9 (0.2)	5 (0.1)	...	...
Other	30 (0.7)	32 (0.7)	...	...
Unclassified	5 (0.1)	7 (0.2)	...	...

(ที่มา: Pedersen TG, et al. JAMA. 2005;294:2437-45.)

- ก. NNT ของ Major coronary event เท่ากับ 87
- ข. NNT ของ Any CHD event เท่ากับ 28
- ค. NNT ของ Fatal or nonfatal stroke เท่ากับ 196
- ง. NNT ของ All-cause mortality เท่ากับ 620
- จ. ถูกทุกข้อ

* หมายเหตุ ผู้ประกอบวิชาชีพเภสัชกรรม สามารถส่งคำตอบมาที่ crm-journal@hotmail.com หรือ http://ccpe.pharmacycouncil.org/index.php?option=article_detail&subpage=article_detail&id=388
โดยระบุ ชื่อ นามสกุล และหมายเลขใบประกอบวิชาชีพ