



บทความการศึกษาต่อเนื่องทางเภสัชศาสตร์ เรื่อง การประเมินความน่าเชื่อถือของงานวิจัยรูปแบบ Randomized Controlled Trial ด้วย CASP Checklists

รหัสการศึกษาต่อเนื่อง: 1010-1-000-001-05-2561

จำนวน: 2.0 หน่วยกิต

วันที่รับรอง: 3 พฤษภาคม 2562

วันที่หมดอายุ: 2 พฤษภาคม 2563

โดย ภก.ดุรงค์กร พลหม

ติดต่อผู้พิมพ์: คณะเภสัชศาสตร์ มหาวิทยาลัยอุบลราชธานี เลขที่ 85 ถ.สกลมารค์ ต.เมืองศรีโค อ.วารินชำราบ
จ.อุบลราชธานี 34190 E-mail: durongkorn.p@ubu.ac.th

วัตถุประสงค์

เพื่อให้ทราบถึงแนวทางการใช้ CASP checklists เป็นเครื่องมือในการประเมินความน่าเชื่อถือของงานวิจัยรูปแบบ randomized controlled trial

บทคัดย่อ

การอ่านและประเมินความน่าเชื่อถือของงานวิจัยประเภท randomized controlled trial เป็นหนึ่งในทักษะที่สำคัญสำหรับเภสัชกรในการแก้ไขปัญหาทางคลินิก ซึ่งปัจจุบันได้มีแนวทางสำหรับการประเมินความน่าเชื่อถือของงานวิจัยจากหลากหลายสถาบัน หนึ่งในนั้นคือ CASP checklists ที่มีการใช้งานอย่างกว้างขวาง อย่างไรก็ตามยังมีประเด็นและรายละเอียดต่าง ๆ ที่ผู้ใช้งานควรทำความเข้าใจ เพื่อให้สามารถใช้อย่างเหมาะสมได้อย่างถูกต้องเหมาะสม และนำไปสู่การตัดสินใจทางคลินิกที่มีคุณภาพ

คำสำคัญ: CASP checklists, randomized controlled trial, critical appraisal

บทนำ

ปัจจุบันโลกได้ก้าวเข้าสู่ยุคสมัยของการแพทย์เชิงประจักษ์ (evidence-based medicine, EBM)¹ การตัดสินใจเพื่อแก้ไขปัญหาทางคลินิกต้องตั้งอยู่บนพื้นฐานของงานวิจัยทางคลินิกที่มีความน่าเชื่อถือ ดังนั้น นอกเหนือจากความรู้ความสามารถทางวิชาการ การอ่านและประเมินความน่าเชื่อถือของงานวิจัยเพื่อนำไปใช้ทางคลินิก (critical appraisal) จึงเป็นอีกทักษะหนึ่งที่สำคัญต่อบุคลากรทางการแพทย์รวมถึงเภสัชกร โดยเฉพาะอย่างยิ่งงานวิจัยในรูปแบบ randomized controlled trial (RCT) ที่พบได้มากในงานวิจัยการแพทย์ เพื่อวิเคราะห์ประสิทธิภาพของวิธีการรักษาที่สนใจเปรียบเทียบกับการรักษาในกลุ่มควบคุม

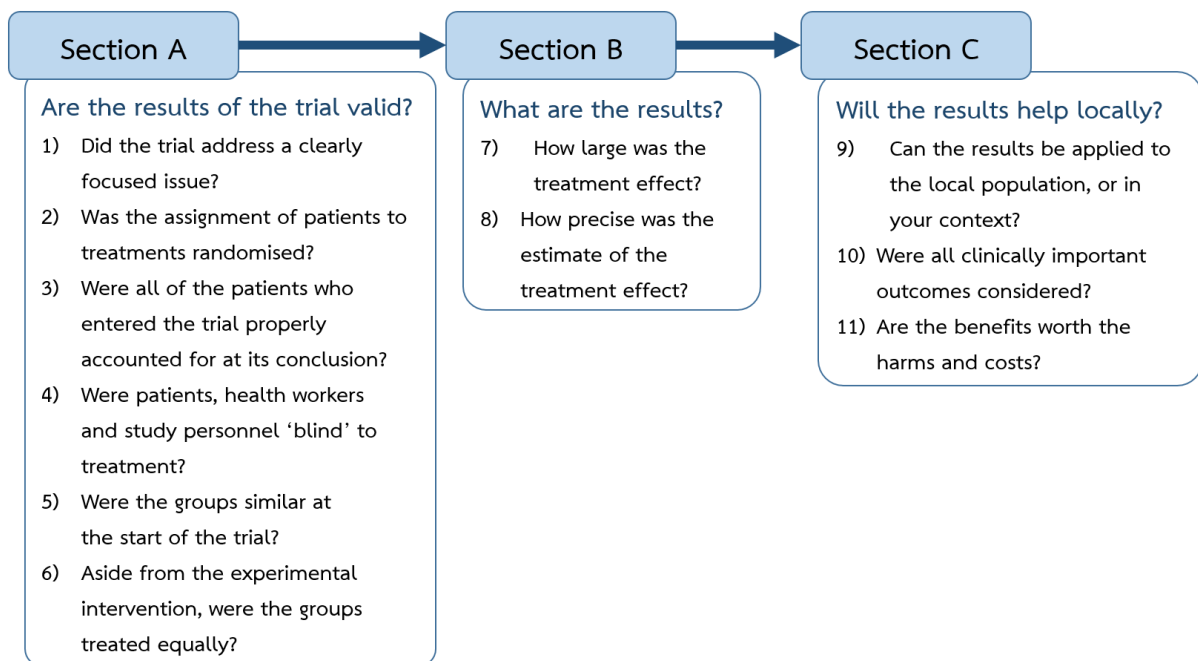
ทั้งนี้ความน่าเชื่อถือของงานวิจัยใด ๆ ขึ้นกับความถูกต้องเหมาะสมของระเบียบวิธีวิจัย เรียกความความเที่ยงตรงภายใน (internal validity) และความถูกต้องเหมาะสมในการนำวิธีทดลองไปประยุกต์ใช้กับ



ผู้ป่วยจริง เรียกความเที่ยงตรงภายนอก (external validity) ซึ่งปัจจุบันได้มีการพัฒนาแนวทางและแบบประเมินความน่าเชื่อถือของงานวิจัยจากหลากหลายสถาบัน หนึ่งในแบบประเมินที่มีการใช้อย่างกว้างขวางคือแบบประเมินของ Critical Appraisal Skills Programme (CASP)²

CASP พัฒนาแบบประเมินความน่าเชื่อถือของงานวิจัยทั้งหมด 8 รูปแบบ โดยงานวิจัยในรูปแบบ RCT นั้นมี CASP randomized controlled trial checklist³ เป็นแบบประเมินที่พัฒนามาจากแนวทางการประเมินงานวิจัยของ JAMA 'Users' guides to the medical literature ปี ค.ศ. 1994 แบบประเมินนี้ประกอบไปด้วยคำถาม 11 คำถาม แบ่งเป็น 3 ส่วน (ดังแผนภาพที่ 1) คือ Section A (ข้อที่ 1 ถึง 6) สำหรับประเมินความเที่ยงตรงภายใน Section B (ข้อที่ 7 และ 8) สำหรับประเมินผลลัพธ์ที่ได้ และ Section C (ข้อที่ 9 ถึง 11) สำหรับประเมินความเหมาะสมในการนำวิธีทดลองไปประยุกต์ใช้กับผู้ป่วย ในแต่ละข้อจะมีคำใบ้ (hint) ช่วยให้คำแนะนำเพิ่มเติมว่าในคำถามนั้นมีประเด็นใดบ้างที่ต้องพิจารณา คำถามแต่ละข้อสามารถตอบได้ 3 แบบคือ ใช่ ไม่ใช่ และไม่สามารถบอกได้

สิ่งหนึ่งที่ต้องทำความเข้าใจก่อนนำไปใช้คือ CASP Checklists ไม่ใช่เครื่องมือสำหรับให้คะแนนงานวิจัย จึงไม่มีคะแนนให้ในแต่ละข้อคำถาม แต่ถูกออกแบบมาเป็นแนวทางในการวิเคราะห์ความน่าเชื่อถือและความเหมาะสมของงานวิจัยเพื่อนำไปประยุกต์ใช้ให้เกิดประโยชน์ต่อผู้ป่วยด้วยดุลยพินิจของผู้ประเมิน



แผนภาพที่ 1 โครงสร้างแบบประเมิน CASP randomized controlled trial checklist

Section A: Are the results of the trial valid?

ในส่วนแรกจะเป็นคำถามเพื่อประเมินความเที่ยงตรงภายในของงานวิจัย โดยพิจารณาจากความถูกต้องเหมาะสมของขั้นตอนตั้งแต่กระบวนการออกแบบวิธีวิจัยไปจนถึงการได้มาซึ่งผลลัพธ์ ประกอบไปด้วยคำถาม 6 ข้อ ได้แก่



1. Did the trial address a clearly focused issue?

HINT: An issue can be ‘focused’ In terms of

- the population studied
- the intervention given
- the comparator given
- the outcomes considered

ในข้อแรกเป็นการถามถึงความชัดเจนของจุดมุ่งหมายหรือคำถามในงานวิจัย โดยพิจารณาได้จากองค์ประกอบหลักที่เรียกว่า PICO^{4,5} ซึ่งประกอบไปด้วย

- 1) Patient หรือ Participant คือลักษณะของผู้ป่วยหรือกลุ่มตัวอย่างที่เข้าร่วมงานวิจัย
- 2) Intervention คือสิ่งที่ทดลอง เป็นสิ่งที่สนใจศึกษาและต้องการทราบถึงผลลัพธ์ที่เกิดขึ้นจากสิ่งนั้น โดยกำหนดให้ใช้ในกลุ่มทดลอง ในงานวิจัยทางการแพทย์ intervention จะเกี่ยวข้องกับวิธีการวินิจฉัย หรือรักษาโรค เช่น ยาที่พัฒนาขึ้นใหม่ หรือแนวทางการรักษาใหม่ ๆ เป็นต้น
- 3) Comparison หรือ Control คือสิ่งที่ให้ในกลุ่มควบคุมเพื่อใช้เปรียบเทียบกับผลลัพธ์ที่ได้จากสิ่งทดลอง เช่น งานวิจัยที่ต้องการทราบถึงประสิทธิภาพของยาชนิดใหม่ ในกลุ่มควบคุมจึงอาจกำหนดให้ใช้ยาตามแนวทางการรักษาที่เป็นมาตรฐานซึ่งมีผลต่อการรักษาจริง (active control) หรืออาจกำหนดให้เป็นยาหลอก (placebo) ซึ่งไม่ได้มีผลในการรักษา
- 4) Outcomes ผลลัพธ์ที่ได้จากสิ่งทดลอง ในทางการแพทย์จะเน้นไปที่อิทธิพลของการรักษา (treatment effect) เช่น การหายขาดจากโรค อาการเจ็บป่วยที่ดีขึ้น หรือผลทางห้องปฏิบัติการที่เกี่ยวข้องกับโรคนั้น เป็นต้น โดยวิเคราะห์เปรียบเทียบความแตกต่างของผลลัพธ์ที่ได้ระหว่างกลุ่มทดลองและกลุ่มควบคุม ทั้งระดับนัยสำคัญทางสถิติ (statistical significance) และนัยสำคัญทางคลินิก (clinical significance)

งานวิจัยที่ระบุองค์ประกอบทั้ง 4 ส่วนของ PICO ไว้อย่างชัดเจน แสดงถึงการมีคำถามของงานวิจัยที่ดี ช่วยให้ผู้อ่านสามารถรับรู้ได้ว่างานวิจัยชิ้นนั้นมีความสอดคล้องกับประเด็นปัญหาที่กำลังพบเจออยู่หรือไม่และง่ายต่อการนำไปประยุกต์ใช้เมื่อประเมินแล้วว่ามีความน่าเชื่อถือเพียงพอ

2. Was the assignment of patients to treatments randomised?

HINT: Consider

- how this was carried out
- was the allocation sequence concealed from researchers and patients

สิ่งที่ต้องพิจารณาในข้อที่สองคือวิธีการที่ผู้วิจัยใช้แบ่งกลุ่มตัวอย่างที่คัดเลือกเข้ามาในงานวิจัยออกเป็นกลุ่มทดลองและกลุ่มควบคุม วิธีการที่ดีคือใช้การสุ่ม (randomization)^{4,5} แบ่งกลุ่มตัวอย่าง เพราะเป็นวิธีที่ทำให้กลุ่มตัวอย่างทุกคนมีโอกาสได้รับรักษาที่เป็นสิ่งทดลองหรือสิ่งเปรียบเทียบเท่า ๆ กัน ทำให้มั่นใจได้ว่าการมีขึ้นของกลุ่มทดลองและกลุ่มควบคุม ไม่ได้เกิดมาจากอคติ (bias) ของผู้วิจัย นอกจากนี้กระบวนการสุ่มที่ดีจะช่วยให้ทั้งกลุ่มทดลองและกลุ่มควบคุมมีคุณลักษณะต่าง ๆ ที่คล้ายคลึงกัน ณ จุดเริ่มต้นของงานวิจัย โดยปกติแล้วกลุ่มตัวอย่างแต่ละคนจะมีปัจจัยต่าง ๆ ทั้งที่รู้และไม่รู้ว่าจะส่งผลต่อผลลัพธ์ของงานวิจัยอย่างการเกิดโรคหรือผลการรักษาหรือไม่ การสุ่มที่ดีจะช่วยให้มีการกระจายตัวของคุณลักษณะเหล่านี้เท่า ๆ กันในกลุ่มทดลองและกลุ่มควบคุม ดังนั้นเมื่อการทดลองดำเนินไปจนได้ผลลัพธ์เกิดขึ้น จะทำให้มั่นใจได้มากขึ้นว่าผลลัพธ์ที่ได้



เกิดจากยาที่ให้ ไม่ใช่เพราะความแตกต่างของคุณลักษณะในกลุ่มตัวอย่างที่แตกต่างกัน ข้อมูลส่วนที่จะบอกถึงลักษณะของกลุ่มตัวอย่างที่ได้หลังจากสุ่มมักจะแสดงอยู่ในส่วนแรกของผลลัพธ์เป็นตารางที่เรียกว่า ตารางคุณลักษณะกลุ่มตัวอย่าง (Baseline Characteristics) ซึ่งจะกล่าวถึงต่อไป

การปกปิดกระบวนการสุ่ม (concealment) เป็นอีกประเด็นหนึ่งที่ต้องพิจารณาถึงเมื่อพูดถึงการสุ่ม อาจทำได้โดยการกำหนดให้ผู้ที่ทำหน้าที่สุ่มกลุ่มตัวอย่างเป็นทีมงานที่ไม่เกี่ยวข้องกับการคัดเลือกกลุ่มตัวอย่างหรือการรักษาโดยตรง เช่น ในงานวิจัยที่แพทย์เป็นผู้คัดเลือกผู้ป่วยเพื่อเข้ารับงานวิจัย อาจกำหนดให้เภสัชกรเป็นผู้ทำการสุ่ม รวมทั้งหลีกเลี่ยงการเผชิญหน้ากันตรง ๆ ระหว่างผู้วิจัยและผู้สุ่ม โดยการใช้การติดต่อผ่านโทรศัพท์หรืออินเทอร์เน็ต สำหรับวิธีการสุ่มนั้นมีหลายวิธี เช่น การใช้ตารางสุ่ม การใช้รหัสสุ่มใส่ซองที่ปิดผนึก หรือการสุ่มด้วยระบบคอมพิวเตอร์ เป็นต้น

3. Were all of the patients who entered the trial properly accounted for at its conclusion?

HINT: Consider

- was the trial stopped early
- were patients analysed in the groups to which they were randomized

ในช่วงระหว่างดำเนินงาน หากกลุ่มทดลองและกลุ่มควบคุมไม่มีการขอลออกจากงานวิจัย ไม่ขาดหายจากการติดตามผล ไม่มีการเปลี่ยนแปลง สลับ หรือย้ายกลุ่มการรักษาจากกลุ่มหนึ่งไปอีกกลุ่มหนึ่ง งานวิจัยที่ได้ย่อมมีความน่าเชื่อถือเพราะความเท่าเทียมของคุณลักษณะภายในกลุ่มตัวอย่างแต่ละกลุ่มจะยังคงเดิมตามที่มีการสุ่มไว้ และยังคงรักษาอำนาจทางสถิติ (power) ไว้ได้

อย่างไรก็ตามงานวิจัยทางคลินิกย่อมมีโอกาสที่จะเกิดปัญหาต่าง ๆ ในข้างต้นได้ ผู้อ่านจึงจำเป็นต้องพิจารณาว่ามีการระบุถึงปัญหาเหล่านี้ไว้หรือไม่ เช่น งานวิจัยดำเนินไปตามระยะเวลาที่กำหนดจริง มีการระบุถึงสัดส่วนกลุ่มตัวอย่างที่ขาดการติดตามผลการรักษา รวมถึงการระบุถึงเหตุผลที่กลุ่มตัวอย่างนั้นขาดการติดตามไปหรือไม่ เช่น ขาดหายไปเพราะหายดีแล้วจึงไม่คิดว่าจำเป็นต้องใช้ยาอีก หรือเพราะอาการแย่ลงเกิดผลข้างเคียงในการรักษาจนทนไม่ไหวทำให้หยุดการรักษา และพิจารณาต่อว่าสัดส่วนที่ขาดหายไปนั้นสูงมากกว่าอีกกลุ่มเกินไปหรือไม่ โดยเฉพาะในกรณีที่ผลลัพธ์ของการวิจัยที่ได้มีจำนวนการปรากฏที่ค่อนข้างน้อย เช่น การเกิดโรคที่มีอุบัติการณ์ต่ำหรือการเสียชีวิตจากโรค การสูญเสียกลุ่มตัวอย่างไปเพียงย่อมส่งผลให้เกิดอคติในการแปลผลลัพธ์ของงานวิจัย กระบวนการที่ผู้วิจัยใช้ในการวิเคราะห์ผลลัพธ์จึงมีความสำคัญ หากผู้วิจัยเลือกวิเคราะห์แบบ per-protocol analysis โดยการตัดข้อมูลจากกลุ่มตัวอย่างที่ขาดการติดตามออก คัดเอาเฉพาะข้อมูลที่ได้จากกลุ่มตัวอย่างที่อยู่ครบตามกระบวนการงานวิจัย วิธีการนี้จะทำให้คุณลักษณะภายในกลุ่มตัวอย่างที่ตั้งใจให้เท่าเทียมกันด้วยการสุ่มเมื่อเริ่มต้นสูญเสียไป ในขณะที่การวิเคราะห์แบบ intention-to-treat analysis ซึ่งเป็นการวิเคราะห์ผลลัพธ์ของงานวิจัยตามกลุ่มที่ได้รับการสุ่มไว้ตั้งแต่แรกเริ่มวิจัย จึงช่วยรักษาอิทธิพลของการสุ่ม คงความเท่าเทียมของคุณลักษณะภายในกลุ่มตัวอย่างไว้ได้ intention-to-treat analysis จึงเป็นวิธีการวิเคราะห์ที่เหมาะสม

Section A ทั้งสามข้อแรกนั้นถูกออกแบบให้เป็นคำถามคัดกรอง (screening questions) ที่สามารถอ่านและหาคำตอบได้อย่างรวดเร็วจากชื่อเรื่อง บทคัดย่อ (abstract) และระเบียบวิธีวิจัย (method) โดยสรุปคือหากสามารถระบุคำถามของงานวิจัยได้และตรงกับประเด็นปัญหาที่ต้องการแก้ไข มีกระบวนการสุ่มกลุ่ม



ตัวอย่างที่ดี กลุ่มตัวอย่างอยู่จนสิ้นสุดงานวิจัย หรือระบุสัดส่วนและสาเหตุของคนที่ย้ายไปอย่างชัดเจนไม่ปกปิด และมีการใช้วิธีการวิเคราะห์ผลที่เหมาะสม จึงค่อยประเมินต่อด้วยคำถามชุดต่อไป

4. Were patients, health workers and study personnel 'blind' to treatment?

นอกจากการปกปิดกระบวนการสุ่ม กระบวนการที่เกิดขึ้นหลังจากนั้นก็ควรต้องได้รับการปกปิดด้วยการปกปิด (blinding)^{4,5} เป็นกระบวนการป้องกันไม่ให้เกิดการระบุตัวตนของกลุ่มตัวอย่างแต่ละคนได้ว่าอยู่ในกลุ่มทดลองหรือกลุ่มควบคุมจากทั้งฝ่ายผู้เก็บข้อมูลและตัวผู้ป่วยเอง เพื่อลดอคติของการเก็บข้อมูล (information bias) เพราะข้อมูลบางชนิดสามารถเปลี่ยนแปลงได้ขึ้นกับการตัดสินใจหรือการรับรู้ของคน เช่น การประเมินผลการรักษา การวิเคราะห์หาสาเหตุของการเสียชีวิตซึ่งต้องอาศัยดุลยพินิจของผู้เก็บข้อมูล หรือความรู้สึกเจ็บปวดซึ่งเป็นสิ่งที่ผู้ป่วยรายงานด้วยความรู้สึกของตนเอง

การปกปิดข้อมูลในฝ่ายของกลุ่มตัวอย่างหรือผู้ป่วย มีขึ้นเพื่อป้องกันอคติที่อาจเกิดขึ้นจากการรับรู้การรักษาที่ตนเองและผู้อื่นได้รับ เช่น ถ้าผู้ป่วยรู้ว่าตนอยู่ในกลุ่มทดลอง ได้รับยาตัวใหม่ซึ่งมีราคาแพงและคาดว่าจะมีประสิทธิผลดีกว่ายาเดิม ความรู้สึกเชื่อมั่นในยาใหม่ที่ได้รับอาจทำให้ผู้ป่วยรายงานผลลัพธ์ประเภที่เป็นความรู้สึกอย่างความเจ็บปวดหรือความพึงพอใจในทางที่ดีมาก นอกจากนี้ยังอาจส่งผลไปถึงพฤติกรรมอื่น ๆ เช่น ความสม่ำเสมอในการใช้ยา การเลือกทานอาหาร การออกกำลังกายที่มากขึ้นหรือน้อยลงกว่าปกติ การเปลี่ยนแปลงเหล่านี้ย่อมส่งผลกระทบต่อการรักษา ทั้งหมดจึงเป็นเหตุผลที่ควรมีการปกปิดข้อมูลจากฝ่ายกลุ่มตัวอย่าง

การปกปิดข้อมูลในฝ่ายของผู้วิจัยหรือผู้เก็บข้อมูล จะช่วยป้องกันอคติจากการเลือกเก็บข้อมูลได้ เช่น หากผู้เก็บข้อมูลเชื่อว่ายาที่ใช้ทดลองมีประสิทธิผลดีกว่ายาเก่า และรู้ว่าผู้ป่วยคนใดอยู่ในกลุ่มทดลอง เมื่อถึงเวลาเก็บข้อมูล ผู้เก็บข้อมูลอาจเลือกเก็บข้อมูลที่แสดงถึงประสิทธิผลที่ดีจากกลุ่มทดลองมากกว่าอีกกลุ่มได้ การปกปิดไม่ให้ผู้เก็บข้อมูลระบุได้ว่าผู้ป่วยแต่ละคนนั้นมาจากกลุ่มใดโดยอมลตปัญหาในส่วนนี้

งานวิจัยที่มีการปกปิดข้อมูลทั้งในฝ่ายของกลุ่มตัวอย่างและผู้เก็บข้อมูล เรียกว่าการ double blind ซึ่งจะช่วยลดอคติจากทั้งสองฝ่ายได้ดังที่อธิบายในข้างต้น ยิ่งถ้ามีการปกปิดในฝ่ายผู้วิเคราะห์ข้อมูลด้วยจะยิ่งทำให้ผลงานวิจัยน่าเชื่อถือขึ้น อย่างไรก็ตามงานวิจัยบางประเภทก็ไม่สามารถปกปิดข้อมูลได้ทั้งหมด เช่น งานวิจัยที่ศึกษาผลของการออกกำลังกายต่อการไม่ออกกำลังกาย กลุ่มตัวอย่างย่อมรู้ได้ด้วยตัวเองว่าอยู่ในกลุ่มไหน หรืองานวิจัยที่ทำในโรคหรือยาที่มีความเสี่ยงต่อชีวิตสูง แพทย์จำเป็นต้องรู้ว่าผู้ป่วยได้รับยาใดอยู่เพื่อที่จะได้ให้การรักษาได้ทันท่วงทีหากเกิดอันตรายที่รุนแรง งานวิจัยในลักษณะนี้อาจปกปิดได้เฉพาะฝ่ายใดฝ่ายหนึ่ง เรียกว่า single blind ผู้อ่านจึงต้องวิเคราะห์ว่า ในงานวิจัยที่อ่านมีการปกปิดหรือไม่ ปกปิดด้วยวิธีใด และทำในฝ่ายใดบ้าง หากมีฝ่ายที่ไม่ถูกปกปิด การรับรู้ของคนในฝ่ายนั้นได้ส่งผลอย่างไรต่อผลลัพธ์ของงานวิจัย ทั้งนี้รายละเอียดของกระบวนการปกปิดข้อมูลสามารถหาได้จากส่วนของระเบียบวิธีวิจัย

5. Were the groups similar at the start of the trial?

HINT: Consider

- other factors that might affect the outcome, such as; age, sex, social class

แม้การสุ่มเป็นวิธีการที่ช่วยลดอคติจากความแตกต่างของปัจจัยที่แตกต่างกันระหว่างกลุ่มทดลองและกลุ่มควบคุม อย่างไรก็ตามยังมีความเป็นไปได้ที่สุดท้ายแล้วคุณลักษณะของกลุ่มตัวอย่างจะมีความแตกต่างกันด้วยความบังเอิญ โดยพิจารณาจากตารางคุณลักษณะกลุ่มตัวอย่าง ที่แสดงข้อมูลเป็นจำนวนหรือสัดส่วนของคุณลักษณะต่าง ๆ ทั้งในกลุ่มตัวอย่างและกลุ่มควบคุมตามผู้วิจัยได้เก็บรวบรวมไว้ รวมทั้งแสดงผลการวิเคราะห์ความแตกต่างทางสถิติของคุณลักษณะระหว่างกลุ่มตัวอย่างเป็นค่า p-value อย่างไรก็ตามในปัจจุบัน



หลาย ๆ งานวิจัยจะไม่แสดงค่า p-value ในตารางคุณลักษณะกลุ่มตัวอย่างให้เห็น ยกเว้นพบว่ามีคุณลักษณะข้อใดที่แตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างกลุ่ม

สิ่งที่ผู้อ่านจะต้องทำการวิเคราะห์จากตารางคุณลักษณะกลุ่มตัวอย่าง คือการค้นหาว่าปัจจัยใดที่มีจำนวนหรือสัดส่วนแตกต่างกันระหว่างสองกลุ่มอย่างมีนัยสำคัญทางสถิติ หากคุณลักษณะของกลุ่มตัวอย่างในงานวิจัยไม่แตกต่างกัน จะสามารถบอกได้ว่างานวิจัยชิ้นนี้มีกระบวนการสุ่มที่ดี แต่สิ่งที่ต้องวิเคราะห์ต่อคือยังคงมีปัจจัยอื่นที่สำคัญต่อผลลัพธ์งานวิจัยที่ผู้วิจัยไม่ได้เก็บหรือแสดงไว้ในตารางคุณลักษณะกลุ่มตัวอย่าง เช่น งานวิจัยเกี่ยวกับความเสี่ยงในการเกิดโรคหลอดเลือดสมองอุดตัน (stroke) ในงานวิจัยนั้นควรแสดงข้อมูลอายุ เพศ เชื้อชาติ โรคความดันโลหิตสูง โรคไขมันในเลือดสูง โรคเบาหวาน โรคหัวใจและหลอดเลือด ประวัติโรคหลอดเลือดสมองและหัวใจในครอบครัว การสูบบุหรี่ และการไม่ออกกำลังกาย ซึ่งเป็นปัจจัยเสี่ยงที่สำคัญในโรคนี้⁶ เพราะปัจจัยที่ไม่ได้เก็บมาอาจเป็นปัจจัยที่มีความแตกต่างกันและอาจส่งผลต่อผลลัพธ์ของงานวิจัย ในกรณีที่พบว่าไม่มีปัจจัยใดที่มีความแตกต่างระหว่างสองกลุ่ม ต้องประเมินต่อว่าปัจจัยดังกล่าวมีโอกาสส่งผลกระทบต่อผลลัพธ์ของงานวิจัยหรือไม่ และผู้วิจัยได้นำเอาจัดการกับปัจจัยนั้นอย่างไร

6. Aside from the experimental intervention, were the groups treated equally?

ในโรคบางโรค เช่น โรคเรื้อรังต่าง ๆ จะต้องมีการใช้ยาหลายชนิดในการรักษา ดังนั้นในงานวิจัยที่เกี่ยวข้องกับโรคที่ต้องได้รับการรักษาหลายชนิด ผู้วิจัยจะต้องมีการควบคุมให้การรักษาอื่น ๆ นอกเหนือจากยาที่ใช้ทดลองและเปรียบเทียบได้รับเหมือนกัน มีการนัดตรวจติดตามด้วยมาตรฐานเดียวกัน เป็นต้น และควรแสดงไว้ในส่วนของระเบียบวิธีวิจัย

Section B: What are the results?

ในส่วนนี้เป็นการประเมินถึงประโยชน์และความแม่นยำผลลัพธ์ที่ได้จากงานวิจัย

7. How large was the treatment effect?

HINT: Consider

- what outcomes were measured
- is the primary outcome clearly specified?
- what results were found for each outcome

ในการประเมินผลลัพธ์ที่เกิดขึ้นจากสิ่งทดลอง ควรเริ่มจากการประเมินผลลัพธ์หลัก^{7,8} (primary outcome) เนื่องจากเป็นตัวแปรสำคัญในการคำนวณขนาดตัวอย่าง (sample size) และอำนาจทางสถิติ ทั้งนี้ผลลัพธ์หลักอาจมีทั้งประเภทที่วัดผลเพียงสิ่งเดียว หรืออาจเป็นประเภทที่ประกอบไปด้วยผลลัพธ์ย่อย ๆ หลายชนิดรวมกันได้ (composite outcome) เช่น งานวิจัยที่เกี่ยวข้องกับโรคหลอดเลือดหัวใจจะมีการกำหนดผลลัพธ์หลักที่เรียกว่า major adverse cardiovascular events (MACE)⁹ ประกอบไปสามผลลัพธ์ย่อย ได้แก่ โรคกล้ามเนื้อหัวใจขาดเลือด โรคหลอดเลือดสมอง และการเสียชีวิตจากโรคหลอดเลือดหัวใจ

สิ่งที่สำคัญคือผลลัพธ์หลักต้องสอดคล้องและตอบวัตถุประสงค์ของงานวิจัย มีเกณฑ์ที่ใช้ในการวัดผลที่ชัดเจน เป็นรูปธรรมเพื่อลดอคติในการแปลผล ขนาดกลุ่มตัวอย่างและระยะเวลาในการติดตามผลที่เพียงพอ มีการรายงานระดับความร่วมมือในการรักษา นอกจากนี้ผลลัพธ์ที่ได้ต้องมีความสำคัญในทางคลินิกกับผู้ป่วย (patient-oriented outcome) ไม่ใช่เป็นผลที่เก็บมาเพียงเพราะเป็นสิ่งที่วัดค่าได้ (surrogate outcome) เช่น งานวิจัยที่วัดค่าระดับฮอร์โมนในเลือดของผู้ป่วย แม้ว่าผลนั้นจะเป็นรูปธรรมและสัมพันธ์กับยาที่ทำการทดลอง แต่ถ้าไม่มีการติดตามถึงอาการทางคลินิกของผู้ป่วยเลย ย่อมทำให้ผลลัพธ์ที่ได้ไม่ครบถ้วน และต้องระวังในส่วนของ composite outcome ว่าผลลัพธ์แต่ละอย่างที่ว่านั้นสัมพันธ์และจัดกลุ่มเข้ากันได้จริง



ในงานวิจัยยังมีผลลัพธ์รอง (secondary outcome) เพื่อความครบถ้วนเกี่ยวกับสิ่งที่เกิดขึ้นจากสิ่งทดลอง โดยเฉพาะในงานวิจัยทางการแพทย์ที่ต้องมีการประเมินทั้งประสิทธิภาพและความปลอดภัยในการรักษา ผลลัพธ์รองจึงมักเป็นข้อมูลเปรียบเทียบผลทางห้องปฏิบัติการ หรือการเกิดอาการข้างเคียงจากการใช้ยา ผลลัพธ์ที่ได้สามารถนำเสนอได้ในหลายรูปแบบ¹⁰ ทั้งการพรรณนาเป็นจำนวน สัดส่วน ร้อยละ หรือค่าเฉลี่ย และผลลัพธ์ที่สะท้อนอิทธิพลของการรักษา เช่น relative risk (RR), relative risk reduction (RRR), absolute risk reduction (ARR), และ number needed to treat (NNT) เป็นต้น (แสดงในตารางที่ 1) โดยเปรียบเทียบระหว่างกลุ่มตัวอย่างสองกลุ่มโดยกำหนดระดับนัยสำคัญทางสถิติ ($\alpha = 0.05$) หากค่า p-value ที่ได้มีค่าน้อยกว่า 0.05 จะถือว่าผลลัพธ์ที่ได้มีนัยสำคัญทางสถิติ ช่วยให้คาดการณ์ได้ว่าความแตกต่างระหว่างกลุ่มที่เกิดขึ้นในงานวิจัยไม่น่าจะเกิดขึ้นจากความบังเอิญ แต่สิ่งที่ต้องคำนึงถึงเสมอคือนัยสำคัญทางสถิติอาจเกิดขึ้นได้จากตัวแปรกวน (confounding) เพราะขาดการควบคุมที่ดี หรือการมีอคติในงานวิจัย เรียกปัญหาที่เกิดขึ้นนี้ว่า type 1 error หรือ false positive นอกจากนี้ค่า p-value ไม่ได้บ่งบอกถึงอิทธิพลการรักษาที่จะเกิดขึ้นกับผู้ป่วย จึงไม่สามารถตัดสินได้ว่ายาที่มีค่า p-value น้อยกว่าจะดีกว่ายาอีกชนิดเมื่อนำมาใช้ เพราะการที่ p-value สูงกว่าหรือเท่ากับ 0.05 อาจเพราะขาดอำนาจทางสถิติเนื่องจากกลุ่มตัวอย่างมีขนาดเล็ก ระยะเวลาที่ติดตามไม่นานพอ หรืออัตราการเกิดผลลัพธ์ในงานวิจัยต่ำกว่าที่คาดการณ์ไว้ รวมถึงกรณีของผลลัพธ์รองซึ่งไม่ได้คำนวณขนาดกลุ่มตัวอย่างเพื่อเก็บมาตั้งแต่ต้น การไม่พบระดับนัยสำคัญทางสถิติจึงเป็นสิ่งที่เกิดขึ้นได้ การไม่พบระดับนัยสำคัญทางสถิติจากปัญหาอำนาจทางสถิติไม่เพียงพอ เรียกว่า type 2 error หรือ false negative

นัยสำคัญทางคลินิกจึงเป็นสิ่งที่ผู้อ่านต้องประเมินควบคู่ไปกับนัยสำคัญทางสถิติ ซึ่งดูได้จากผลลัพธ์ว่าอัตราการเกิดผลลัพธ์จริงในงานวิจัย (true event rate) เป็นไปตามที่คาดการณ์ไว้ตอนคำนวณกลุ่มตัวอย่างหรือไม่ (expected event rate) การอ่านผลค่า RR ที่บอกถึงโอกาสหรือความเสี่ยงที่เกิดขึ้นในกลุ่มทดลองต่อกลุ่มควบคุม ค่า ARR ที่บอกถึงความแตกต่างสัมบูรณ์ของอัตราการเกิดผลระหว่างสองกลุ่ม ค่า RRR ที่บอกถึงการลดลงของอัตราการเกิดผลลัพธ์ในกลุ่มทดลองต่อกลุ่มควบคุม ค่า NNT ที่บอกจำนวนของผู้ป่วยที่ต้องได้รับการรักษาตามวิธีการที่ใช้ในกลุ่มทดลองเพื่อให้เห็นผลการรักษาในผู้ป่วยหนึ่งคนระดับนัยสำคัญทางคลินิกจะเป็นสิ่งที่ช่วยบอกให้ทราบถึงโอกาสที่จะเกิดขึ้นเมื่อนำวิธีการรักษาในงานวิจัยไปใช้กับผู้ป่วยจริง

8. How precise was the estimate of the treatment effect?

HINT: Consider

- what are the confidence limits

ผลลัพธ์ของงานวิจัยจะเห็นว่ามีการรายงานผลออกมาเป็นค่าคงที่หนึ่งค่าร่วมกับช่วงของตัวเลขชุดหนึ่งด้วยเสมอ เช่น HR=0.89 (95%CI, 0.89-0.98) ช่วงตัวเลขนั้นเรียกว่า ช่วงความเชื่อมั่น^{7,8,10} (confidence interval, CI) ซึ่งมาจากหลักในการประมาณค่าเพื่ออนุมานผลลัพธ์ที่เกิดขึ้นในกลุ่มตัวอย่างไปยังกลุ่มประชากร โดยทั่วไปมักรายงานช่วงความเชื่อมั่นไว้ที่ระดับร้อยละ 95 (95%CI) คือ มั่นใจได้ 95% ว่าช่วงของตัวเลขที่ได้นี้จะครอบคลุมค่าคงที่จริงในประชากร โดยผู้อ่านสามารถประเมินความแม่นยำของผลงานวิจัยได้จากช่วงความ



ประเภทของผลลัพธ์	ความหมาย	ตัวอย่างการแปลผล
Relative Risk (RR)	<p>กลุ่มทดลองมีความเสี่ยงในการเกิดผลลัพธ์เป็นเท่าเมื่อเทียบกับกลุ่มควบคุม</p> <ul style="list-style-type: none"> RR = 1 ทั้งสองกลุ่มมีความเสี่ยงไม่แตกต่างกัน RR < 1 กลุ่มทดลองมีความเสี่ยงในการเกิดผลลัพธ์น้อยกว่ากลุ่มควบคุม RR > 1 กลุ่มทดลองมีความเสี่ยงในการเกิดผลลัพธ์มากกว่ากลุ่มควบคุม 	<p>RR = 0.32 กลุ่มทดลองมีความเสี่ยงในการเกิดผลลัพธ์เป็น 0.32 เท่าของกลุ่มควบคุม หรือการให้สิ่งทดลองช่วยลดความเสี่ยงการเกิดผลลัพธ์ 68% เมื่อเทียบกับสิ่งควบคุม (RRR)</p> <p>RR = 0.67 กลุ่มทดลองมีความเสี่ยงในการเกิดผลลัพธ์เป็น 0.67 เท่าของกลุ่มควบคุม หรือการให้สิ่งทดลองช่วยลดความเสี่ยงการเกิดผลลัพธ์ได้ 33% เมื่อเทียบกับสิ่งควบคุม (RRR)</p>
Relative Risk Reduction (RRR) Relative Risk Increase (RRI)	<p>การรายงานผลโดยปรับจากค่า RR ให้อยู่ในเปอร์เซ็นต์</p> <p>ใช้บอกกว่าสิ่งทดลองสามารถลดหรือเพิ่มผลลัพธ์ได้กี่เปอร์เซ็นต์ เมื่อเทียบกับสิ่งควบคุม (RRR = 1 - RR)</p>	<p>RR = 1.50 กลุ่มทดลองมีความเสี่ยงในการเกิดผลลัพธ์เป็น 1.5 เท่าของกลุ่มควบคุม หรือการให้สิ่งทดลองเพิ่มความเสี่ยงการเกิดผลลัพธ์ 55% เมื่อเทียบกับสิ่งควบคุม (RRI)</p>
Absolute Risk Reduction (ARR) Absolute Risk Increase (ARI)	<p>บอกค่าความแตกต่างสัมบูรณ์ของอัตราการเกิดผลลัพธ์ในกลุ่มควบคุมต่อกลุ่มทดลอง</p> <ul style="list-style-type: none"> ARR = 0 ไม่มีความแตกต่างกันระหว่างสองกลุ่ม ARR < 0 กลุ่มทดลองเกิดผลลัพธ์มากกว่ากลุ่มควบคุม (เรียก ARI) ARR > 0 กลุ่มทดลองเกิดการเกิดผลลัพธ์น้อยกว่ากลุ่มควบคุม 	<p>ARR = 0.04 กลุ่มทดลองเกิดผลลัพธ์น้อยกว่ากลุ่มควบคุม 4%</p> <p>ARR = 0.04 NNT = 1 / 0.04 = 25</p>
Number Needed to Treat (NNT)	<p>จำนวนผู้ป่วยที่ต้องได้รับการรักษาด้วยสิ่งทดลองตามวิธีเนงานวิจัย เพื่อให้เห็นผลการป้องกันรักษาผลลัพธ์ในผู้ป่วยหนึ่งคน (NNT = 1/ARR)</p>	<p>ARR = 0.04 NNT = 1 / 0.04 = 25</p> <p>ต้องให้การรักษาด้วยสิ่งทดลองแก่ผู้ป่วยจำนวน 25 คนจึงจะเห็นผลการรักษาในผู้ป่วยหนึ่งคน</p>

ตารางที่ 1 ผลลัพธ์ที่แสดงอิทธิพลการรักษา¹⁰



เชื่อมั่น การประมาณค่าที่แม่นยำจะได้ช่วงความเชื่อมั่นที่แคบ เปรียบเสมือนการปลูกดอกที่แม่นยำจะทำให้ลูกดอกไปปักอยู่ใกล้ ๆ กันเสมอ อย่างไรก็ตามความกว้างแคบของช่วงความเชื่อมั่นไม่มีเกณฑ์กำหนดตัดสินที่ชัดเจน นอกจากนี้ยังสัมพันธ์กับขนาดของกลุ่มตัวอย่างด้วย ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ มักจะส่งผลให้ได้ช่วงความเชื่อมั่นที่แคบลง เพราะกลุ่มตัวอย่างที่เยอะขึ้น ทำให้มีโอกาสเห็นผลลัพธ์ที่แท้จริงสูงขึ้น ความแม่นยำจึงเพิ่มขึ้น ดังนั้นในการอ่านช่วงความเชื่อมั่นจึงต้องพิจารณาในมุมของอิทธิพลทางคลินิกที่จะเกิดขึ้นด้วย เช่น ในงานวิจัยยาลดไขมันในเลือดตัวใหม่พบว่ามีประสิทธิภาพในการลดระดับไขมันได้ดีกว่ายาเก่า 2.25% (95%CI, 1.55-3.05) แปลว่ายาชนิดใหม่จะสามารถลดระดับไขมันในเลือดได้ดีกว่ายาเก่าประมาณ 1.55 ถึง 3.05% ที่ระดับความเชื่อมั่น 95% สิ่งที่ต้องคิดตามต่อก็คือ การที่ยาตัวใหม่สามารถลดระดับไขมันได้ดีกว่ายาเก่าประมาณ 2-3% นี้ส่งผลต่อระดับไขมันที่ลดลงของผู้ป่วยมากน้อยเพียงใด เป็นประโยชน์ต่ออาการทางคลินิกของผู้ป่วยอย่างไร และคุ้มกับค่าใช้จ่ายที่ต้องเพิ่มขึ้นจากการเปลี่ยนไปใช้ยาใหม่ที่แพงกว่าหรือไม่ เป็นต้น

Section C: Will the results help locally?

ส่วนสุดท้ายเป็นการประเมินความเที่ยงตรงภายนอก (external validity) คือการประเมินความเหมาะสมของนำวิธีในการทดลองไปใช้กับผู้ป่วยจริง ๆ

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

HINT: Consider whether

- the patients covered by the trial are similar enough to the patients to whom you will apply this
- how they differ

พิจารณาว่าผู้ป่วยที่เราต้องการรักษาหรือแก้ปัญหา มีคุณลักษณะที่เหมือนหรือแตกต่างไปจากผู้ป่วยที่คัดเข้ามาในงานวิจัยอย่างไร ถ้าคุณลักษณะมีความแตกต่างกันมาก เราจะไม่สามารถบอกได้ว่าเมื่อให้การรักษาตามแนวทางที่ใช้ในงานวิจัยต่อผู้ป่วยของเราไปแล้วจะได้ผลลัพธ์ในการรักษาไปในแนวเดียวกัน ผู้ป่วยที่เหมาะสมจึงต้องผู้ป่วยมีคุณลักษณะตรงกับเกณฑ์คัดเข้า (inclusion criteria) และไม่มีข้อห้ามใดตามที่กำหนดไว้ในเกณฑ์คัดออก (exclusion criteria) และมีคุณลักษณะโดยรวมที่ใกล้เคียงกับข้อมูลในตารางคุณลักษณะกลุ่มตัวอย่าง

10. Were all clinically important outcomes considered?

HINT: Consider whether

- there is other information you would like to have seen
- if not, does this affect the decision

พิจารณาว่ายังมีข้อมูลทางคลินิกได้อื่นอีกหรือไม่ที่ไม่ปรากฏในงานวิจัย แต่อาจเป็นประเด็นปัญหาในผู้ป่วยรายนี้ได้ เช่น การที่ผู้ป่วยมีโรคร่วมบางอย่างที่ไม่ได้มีการระบุไว้ในงานวิจัย ผู้อ่านย่อมจำเป็นที่จะต้องค้นหาข้อมูลจากแหล่งอ้างอิงที่น่าเชื่อถืออื่น ๆ เพิ่มเติมเพื่อประเมินความปลอดภัยในส่วนนี้ หรือในงานวิจัยต้องมีการตรวจติดตามผลทางห้องปฏิบัติการบางอย่างซึ่งในสถานการณ์นั้นไม่สามารถทำได้ ผู้อ่านจำเป็นจะต้องตัดสินใจว่าจะรับมือกับความเสี่ยงนี้ได้อย่างไร

11. Are the benefits worth the harms and costs?

HINT: Consider

- even if this is not addressed by the trial, what do you think?



ในขั้นสุดท้ายคือการประเมินว่า ในท้ายที่สุดแล้วการนำผลงานวิจัยมาประยุกต์ใช้กับผู้ป่วยนั้นเป็นสิ่งที่มี
จำเป็น ได้ประโยชน์ต่อการรักษา คำนวณเมื่อเทียบกับความเสี่ยงที่ผู้ป่วยจะต้องได้รับ รวมไปถึงผลในด้านอื่น ๆ
ทั้งภาระค่าใช้จ่าย และคุณภาพชีวิตของผู้ป่วยและครอบครัว

บทสรุป

จะเห็นได้ว่า CASP checklists เป็นแบบประเมินที่ออกแบบมาให้มีความกระชับและตรงประเด็นต่อ
การวัดความเที่ยงตรงของงานวิจัยด้วย 11 ข้อคำถาม แต่การที่จะได้มาซึ่งคำตอบในแต่ละข้อคำถามนั้นก็ยังมี
จำเป็นต้องอาศัยทักษะและการฝึกฝนของผู้อ่านในการคิดวิเคราะห์ถึงความน่าเชื่อถือของงานวิจัยทั้งในด้าน
ระเบียบวิธีวิจัยและความเหมาะสมกับปัญหาทางคลินิกที่เผชิญเพื่อนำไปสู่การรักษาและแก้ไขปัญหาของผู้ป่วย
อย่างมีประสิทธิภาพและเกิดประโยชน์สูงสุด

เอกสารอ้างอิง

1. Thoma A. A Brief History of Evidence-Based Medicine (EBM) and the Contributions of Dr David Sackett [Internet]. 2015. Cited 2019 Feb 2. Available at: <https://academic.oup.com/asj/article/35/8/NP261/251339>
2. Critical Appraisal Skills Programme. CASP Bibliography [Internet]. 2018. Cited: 2 Feb 2019. Available at: <https://casp-uk.net/bibliography/>
3. Critical Appraisal Skills Programme. CASP Randomised Controlled Trial Checklist [Internet]. 2018. Cited: 2019 Feb 2. Available at: <https://casp-uk.net/casp-tools-checklists/>
4. Al-Jundi A, Sakka S. Critical Appraisal of Clinical Research. J Clin Diagn Res. 2017; 11(5): JE01-JE05.
5. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem solving. Br Med J. 1995 Apr 29; 310(6987): 1122.
6. Furie K.L., Rost N.S. Overview of secondary prevention of ischemic stroke [Internet]. 2017. Cited: 2019 Feb 2. Available at: https://www.uptodate.com/contents/overview-of-secondary-prevention-of-ischemic-stroke?search=stroke%20risk%20factors&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1
7. CourcierDuplantier S, Falissard B, Fender P. Subjective Outcome Measures of Drug Efficacy. Thérapie. 2003 May 1; 58(3): 26773.
8. Pocock S.J., Stone GW. The Primary Outcome Is Positive — Is That Good Enough? [Internet]. 2016. Cited: 2019 Feb 2. Available at: https://www.nejm.org/doi/full/10.1056/NEJMra1601511?fbclid=IwAR2lWYmRlnpzOFUNRo07c3eO4x0reYz1Y1uZ-AKnY-Wiw_ZV-oilLBUyt44#t=article
9. Kip, K.E., Hollabaugh, K., Marroquin, O.C., Williams, D.O. The Problem with Composite End Points in Cardiovascular Studies: The Story of Major Adverse Cardiac Events and Percutaneous Coronary Intervention. J Am Coll Cardiol. Vol. 51(7): 701-7.
10. Dawson, Beth, Trapp, Robert G. Basic & clinical biostatistics, 4 ed. United States: New York : Lange Medical Books-McGraw-Hill, Medical Pub. Division, c2004.; 2004.

