

Easy Tricks in Article Appraisal

วันชัย เดชสมฤทธิฤทัย พ.บ.

สาขาวิชาโรคระบบการหายใจและวัณโรค
ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล

บทนำ

แพทย์ในยุคปัจจุบันรวมทั้งในอนาคตอยู่ในโลกของข้อมูลข่าวสาร ซึ่งในบริบทของวิชาชีพแพทย์ องค์ความรู้ใหม่ที่เกิดขึ้นจะถูกนำเสนอในรูปแบบของบทความในวารสารวิชาการ เทคโนโลยีสมัยใหม่ทำให้การค้นหาข้อมูลเป็นเรื่องง่ายดาย แต่สิ่งที่ยากและสำคัญสำหรับแพทย์มากกว่าก็คือทักษะในการประเมินความน่าเชื่อถือของข้อมูลที่ได้มา มีบทความตีพิมพ์ที่แนะนำวิธีการดังกล่าวอยู่หลายชิ้น¹ แต่บทความนี้จะนำเสนอวิธีการอย่างง่ายๆ เบื้องต้นที่เป็นพื้นฐานสำหรับการประเมินข้อมูลวิชาการดังกล่าว สำหรับแพทย์ทั่วไปเพื่อใช้ในการฝึกฝนทักษะดังกล่าวต่อไป ในที่นี้จะเน้นหนักสำหรับงานวิจัยในรูปแบบ interventional study ซึ่งเป็นประเภทที่แพร่หลายในปัจจุบัน

เริ่มต้นด้วยคำถาม

Begin with questions

ในการค้นหาข้อมูล แพทย์ควรเริ่มต้นด้วยคำถาม ซึ่งในที่นี้อาจเริ่มจากปัญหาที่แพทย์ประสบในขณะที่ให้การดูแลรักษาผู้ป่วย แนะนำให้แปลงปัญหาที่พบให้เป็นคำถามเชิงวิจัย โดยใช้หลักของ PICO² ได้แก่

1. Population หรือ problem ซึ่งอาจมีความจำเพาะในระดับที่ต่างกัน ขึ้นกับบริบทที่สนใจ เช่น เป็นประชากรที่เป็นโรคนี้โดยทั่วไป หรือเฉพาะกลุ่มที่มีความรุนแรงสูง เป็นต้น

2. Intervention โดยดูว่าการศึกษานั้น ตั้งใจศึกษาวิธีการรักษาที่เราสนใจหรือไม่

3. Comparison โดยประเมินจากบทความว่า มีการเปรียบเทียบกับวิธีการรักษาอื่นหรือไม่ และสิ่งที่นำมาเปรียบเทียบมีความเหมาะสมหรือไม่ โรคใดที่มีการรักษามาตรฐานอยู่แล้วก็ควรทำการศึกษเปรียบเทียบกับวิธีการมาตรฐานแทนที่จะเปรียบเทียบกับ placebo จึงจะถูกต้องทางจริยธรรมการวิจัย

4. Outcome ผลลัพธ์ของการวิจัยที่เป็นเป้าหมายบางอย่าง มีความน่าเชื่อถือในตัวเองสูง เช่น อัตราการตาย เนื่องจากเป็นผลลัพธ์ที่เป็นเป้าหมายในเวชปฏิบัติเช่นกัน แต่การศึกษาถึงอัตราการตายอาจต้องใช้ทรัพยากรมาก จึงอาจต้องใช้ผลลัพธ์อื่นแทน แพทย์จึงควรประเมินว่าผลลัพธ์ที่ตั้งไว้มีความเหมาะสมและน่าเชื่อถือเพียงใด และจะสามารถแปลผลลัพธ์ที่ได้มาใช้ในเวชปฏิบัติได้หรือไม่

ตารางที่ 1. ประเภทของงานวิจัย

Secondary study	Systematic review
Primary study	Randomized controlled trial
	Cohort study
	Case-control study
	Case series
	Case report

ลักษณะเด่นของงานวิจัยแต่ละประเภท

Key features of research method

โดยทั่วไป ความน่าเชื่อถือของงานวิจัย เรียงลำดับตามตารางที่ 1. กล่าวคือ randomized controlled trial จะมีความน่าเชื่อถือมากกว่า cohort study เนื่องจากมีกระบวนการ randomization ทำให้โอกาสที่จะมี confounder น้อย การที่มีความรู้เกี่ยวกับลักษณะเด่นและข้อจำกัดของงานวิจัยแต่ละประเภทจะช่วยให้สามารถประเมินงานวิจัยได้ดีขึ้น อย่างไรก็ตาม อาจต้องพิจารณาองค์ประกอบอื่นๆ ร่วมด้วย เช่น randomized controlled trial ที่ทำในประชากรที่มีจำนวนมากพอก็อาจมีความน่าเชื่อถือมากกว่า systematic review ที่รวบรวมงานวิจัยได้ไม่ครอบคลุมหรือมีความแตกต่างกันมาก; โรคนางโรคมีอุบัติการณ์น้อย การทำการวิจัยอาจทำได้เพียงวิธีการแบบ retrospective; การวิจัยแบบ intervention อาจไม่สามารถทำได้ในบางกรณี เนื่องจากผิดหลักจริยธรรม จึงต้องทำเป็น observational study เป็นต้น อย่างไรก็ตาม ในส่วนของ primary research ถือว่า interventional study ในรูปแบบ randomized controlled trial จะมีความน่าเชื่อถือสูงที่สุด ทั้งนี้เนื่องจากกระบวนการ randomization ที่ถูกวิธีจะกำจัด selection bias และ confounders ที่มีอยู่ออกไปได้ และหากมีกระบวนการ blinding ก็จะสามารถกำจัด bias ที่อาจเกิดขึ้นจากการปฏิบัติที่แตกต่างกันของผู้ที่มีส่วนร่วมในงานวิจัยได้ ดังนั้น การประเมินงานวิจัยควรดูว่าผู้วิจัยได้วางแผน และปฏิบัติได้ถูกต้องตามระเบียบวิธีหรือไม่ โดยเฉพาะกระบวนการ randomization และ blinding ดังกล่าว โดยดูได้ในหัวข้อ Methods ของบทความ

สุ่มจริง/สุ่มหลอก/การปกปิด

True randomization/Pseudo-randomization/Concealment

ในบางครั้งผู้วิจัยอาจใช้วิธีที่คล้ายการสุ่มแต่ในเชิงวิชาการไม่ใช่ true randomization ยกตัวอย่างเช่น การเลือกวิธีการรักษาสลับไปมาตามลำดับที่เข้าร่วมโครงการ หรือเลือกกำหนดตามตัวเลขของวันที่ลงท้ายด้วยเลขคู่หรือคี่

เป็นต้น ซึ่งวิธีการเหล่านี้ไม่จัดว่าเป็น true randomization โดยหลักการแล้ว ผู้ป่วยทุกรายต้องมีโอกาสเท่าๆ กันที่จะถูกเลือกรักษาด้วยวิธีการใด

อีกประการหนึ่ง ผู้มีส่วนร่วมในการวิจัยต้องไม่ทราบล่วงหน้าว่าผู้ป่วยรายต่อไปจะถูกกำหนดการรักษาด้วยวิธีการใด ดังนั้นจึงจะต้องมีวิธีการปกปิดผลการสุ่มเลือก เช่น ใส่ซองปิดผนึกไว้ อย่างไรก็ตามในปัจจุบัน งานวิจัยส่วนใหญ่มักใช้วิธีสุ่มในขณะที่เข้าโครงการ โดยใช้ระบบคอมพิวเตอร์ หรือจากศูนย์ควบคุมกลาง เพื่อหลีกเลี่ยง selection bias ดังกล่าว

ความสำคัญของตารางที่ 1

The importance of Table 1

แม้ว่าผู้วิจัยจะระบุวิธีการของ randomization และ concealment ที่ถูกต้อง แต่อาจไม่ได้ปฏิบัติตามที่ระบุไว้ ข้อมูลจากตารางที่ 1 ของบทความทุกบทความ ซึ่งมักเป็น demographic data ของผู้ป่วยจึงมีความสำคัญอย่างมาก โดยทั่วไป หากกระบวนการของ randomization ถูกต้อง และจำนวนประชากรมากเพียงพอ ลักษณะต่างๆ ของกลุ่มประชากรทั้งกลุ่มศึกษาและกลุ่มควบคุมจะใกล้เคียงกัน ดังนั้นหากพบว่าข้อมูลมีความแตกต่างกัน ให้ตั้งข้อสังเกตว่า อาจมีความคลาดเคลื่อนในกระบวนการของ randomization ยิ่งมีความแตกต่างกันหลายข้อก็ยิ่งน่าสงสัยมากขึ้น ในบางบทความ ผู้วิจัยอาจทำการเปรียบเทียบข้อมูลของทั้ง 2 กลุ่มด้วยวิธีทางสถิติ และแสดง p-value มาให้ด้วย อย่างไรก็ตาม อาจไม่มีความจำเป็นต้องแสดง p-value เนื่องจากจำนวนประชากรอาจไม่มากพอที่จะแสดงถึงความแตกต่างกัน

ปัจจัยซ่อนเร้น

Find the hidden ones

งานวิจัยโดยทั่วไปมีจุดประสงค์เพื่อทดสอบสมมติฐานว่ามีความสัมพันธ์ระหว่าง exposure กับ outcome หรือไม่ แต่ outcome ที่เกิดขึ้นอาจจะเกิดจากปัจจัยอื่นที่นอกเหนือจาก exposure ที่เราต้องการศึกษาก็ได้ (เราเรียกปัจจัยนี้ว่า confounder) ในทางระบาดวิทยามีวิธีจัดการกับ confounder ดังต่อไปนี้

1) กรณีที่เป็น observational study มีวิธีจัดการกับปัญหา confounder โดยการเก็บข้อมูลที่เกี่ยวข้องกับ confounder ร่วมด้วยตั้งแต่แรก แล้วนำข้อมูลเหล่านี้มาวิเคราะห์ทางสถิติในภายหลัง ที่เรียกว่าวิธี multivariate analysis ดังนั้น ควรประเมินว่าผู้วิจัยได้ระบุถึงปัจจัยที่อาจจะเป็น confounder ได้ครบถ้วนตั้งแต่แรกและเก็บข้อมูลไว้หรือไม่ หากเก็บข้อมูลครบถ้วน และนำมาวิเคราะห์ด้วยวิธี multivariate analysis ก็ถือว่าผู้วิจัยได้ปฏิบัติเต็มที่แล้ว อย่างไรก็ตาม ยังอาจมี confounder ซึ่งยังไม่เป็นที่รู้จักมาก่อนหลงเหลืออยู่ (residual confounder) ซึ่งถือเป็นข้อจำกัดของ observational study

2) กรณีของ randomized controlled trial มีข้อได้เปรียบคือ กระบวนการ randomization จะทำให้ปัจจัยที่อาจเป็น confounder (รวมทั้ง residual confounder) ถูกกระจายไปทั่วกลุ่มศึกษาและควบคุม จึงไม่ส่งผลต่อ outcome ของการศึกษาแต่ประการใด

ขนาดนั้นสำคัญไหม

The matter of size

จะประเมินอย่างไรว่างานวิจัยนั้นมีจำนวนประชากรเพียงพอ เบื้องต้นให้ดูจากวิธีการคำนวณ sample size ว่าถูกต้องหรือไม่ โดยทั่วไปมักใช้ค่า α error = 0.05, power = 0.8-0.9, และค่า outcome difference อ้างอิงจากการศึกษาก่อนหน้า ซึ่งการศึกษาส่วนใหญ่มักจะได้จำนวนประชากรที่ใกล้เคียงกับค่าที่คำนวณไว้ กรณีที่ผลการศึกษาเป็น positive study ($p < 0.05$) ให้ถือว่าจำนวนประชากรเพียงพอ แต่หากผลเป็น negative study ควรย้อนกลับมาดูว่าจำนวนประชากรน้อยกว่าจำนวนที่คำนวณไว้หรือไม่ ถ้าน้อยกว่าพอสมควร เป็นไปได้ว่า ผล negative นั้นอาจเกิดจากจำนวนประชากรที่ไม่เพียงพอ อย่างไรก็ตาม อาจมีกรณีที่อัตราการเกิด outcome ของการศึกษานี้แตกต่างจากการศึกษาก่อนหน้าที่นำมาใช้คำนวณ อาจใช้วิธีคำนวณย้อนกลับเพื่อดูว่า power เพียงพอหรือไม่

รูปที่ 1 ก็สำคัญ

Figure 1 is also important

นอกเหนือจากตารางที่ 1 ที่ช่วยบอกถึงกระบวนการ randomization แล้ว รูปที่ 1 ของบทความซึ่งมักจะเป็น diagram แสดงขั้นตอน enrollment, randomization ไปจนถึง outcome ของการศึกษา และบอกถึงจำนวนของผู้ป่วยในแต่ละขั้นตอน ก็มีความสำคัญเช่นกัน ในระหว่างที่งานวิจัยดำเนินไป จะมีผู้ป่วยจำนวนหนึ่งต้องออกจากการศึกษาก่อนที่จะสิ้นสุดโครงการ ซึ่งเกิดได้จากสาเหตุหลายประการ ทั้งนี้ หากมีประชากรโดยรวมที่ขาดหายไปเกิน 20% พึงระวังว่า อาจทำให้ผลการศึกษาลดเคลื่อนจากความเป็นจริงจาก selection bias ได้ หรือหากพบว่าสัดส่วนของประชากรที่ขาดไปมีความแตกต่างกันมากระหว่างกลุ่ม ซึ่งอาจเกิดจากปัจจัยบางอย่างและทำให้เกิดอคติได้เช่นกัน กรณีนี้ผู้วิจัยอาจแก้ไขโดยการทำการวิเคราะห์แบบ intention to treat เพื่อลดอคติที่อาจเกิดขึ้น

สถิติคือยาขม

Statistical mistakes

สถิติเป็นเรื่องเข้าใจยาก แต่ก็มีหลักพื้นฐานบางอย่างที่ควรรู้ ข้อผิดพลาดทางสถิติที่อาจพบในบทความ เช่น ในกรณีที่มีความแตกต่างของปัจจัยบางอย่างระหว่างกลุ่มควบคุมกับกลุ่มศึกษาในเบื้องต้น หากผู้วิจัยไม่ได้ใช้วิธีการทางสถิติทำการวิเคราะห์และปรับ (adjust) อาจทำให้ปัจจัยบางอย่างยังคงเป็น confounder ต่อผลลัพธ์ของการศึกษาได้ นอกจากนั้น ผู้วิจัยอาจเลือกใช้วิธีทางสถิติที่ไม่เหมาะสม เช่น ใช้วิธี parametric analysis สำหรับข้อมูลที่ไม่ได้กระจายตัวแบบ Normal เนื่องจากไม่ได้ทดสอบการกระจายตัวของข้อมูลก่อน หรือมีการรายงานค่าเฉลี่ยที่ผิดประเภท เช่น รายงานเป็นค่า mean สำหรับข้อมูลที่ไม่ได้กระจายแบบ Normal แทนที่จะรายงานเป็น median เป็นต้น ในบางครั้ง ผลการศึกษาเป็น negative สำหรับ main outcome ผู้วิจัยจึงพยายามวิเคราะห์ปัจจัยอื่นๆ เพิ่มเติมภายหลัง โดยไม่ได้มีการวางแผนไว้แต่แรก ซึ่งก็อาจมีบางปัจจัยที่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ เรียกว่าวิธีนี้ว่า post hoc study ในกรณีนี้ ความน่าเชื่อถือจะต่ำ เนื่องจากอาจเป็นความแตกต่างที่เกิดโดยบังเอิญได้

สัมพันธภาพเพียงใด

Cause-effect relationship

ผลการวิเคราะห์ทางสถิติที่พบว่ามี ความแตกต่างอย่างมีนัยสำคัญ ($p < 0.05$) นั้นบอกเพียงว่ามีความสัมพันธ์กันของ 2 ปัจจัยนั้น แต่ไม่ได้แปลว่าปัจจัยหนึ่งจะเป็นสาเหตุของอีกปัจจัยหนึ่งเสมอไป การที่จะบ่งบอกถึง cause-effect relationship นั้น อาศัยการพิจารณาองค์ประกอบ (Bradford Hill's criteria) ต่อไปนี้ (ตารางที่ 2)³

ตารางที่ 2. Bradford-Hill's criteria³

เกณฑ์	ลักษณะของ cause-effect relationship
Strength of association	ขนาดของ effect size มาก
Biologic plausibility	สามารถอธิบายได้ด้วยกลไกทางชีววิทยา
Consistency	ผลการศึกษาหลายการศึกษาไปในทิศทางเดียวกัน
Temporality	ปัจจัยที่เป็นสาเหตุควรมาก่อนผล
Dose-response relationship	ขนาดของผลลัพธ์เป็นไปตามขนาดของปัจจัยต้นเหตุ

นัยสำคัญแบบไหน

Statistical v Clinical significance

การพบว่าการรักษาใดดีกว่ากลุ่มควบคุมโดยมีนัยสำคัญทางสถิติอย่างเดียว ไม่เพียงพอต่อการตัดสินใจนำไปใช้ในเวชปฏิบัติ ทั้งนี้เนื่องจากความแตกต่างเพียงเล็กน้อยอาจมีนัยสำคัญทางสถิติได้หากใช้ sample size ที่มากเกินไป ควรพิจารณาว่าความแตกต่างระดับนั้นมีความสำคัญทางคลินิกหรือไม่ ปัจจุบันมีการกำหนดค่า minimal clinically important difference สามารถค้นหาประกอบ การตัดสินใจได้ นอกจากนี้ ควรพิจารณาค่า absolute risk reduction ร่วมด้วย ซึ่งสามารถเข้าใจได้ง่ายขึ้น โดยการแปลง

เป็นค่า number needed to treat (NNT) ถ้ามีค่าน้อยยิ่งบ่งบอกถึงความสำคัญทางคลินิกที่มาก

ความสำคัญของบริบท

Real life

การนำผลการศึกษามาใช้ในเวชปฏิบัติ ต้องคำนึงถึงปัจจัยบางอย่างเพิ่มเติม หากงานวิจัยทำการศึกษาในประชากรที่มีบริบทที่แตกต่าง เช่น ชาติพันธุ์ ภูมิประเทศ ภูมิอากาศ ปัจจัยทางสังคมและเศรษฐกิจ เป็นต้น ควรพิจารณาว่าปัจจัยเหล่านี้มีโอกาสทำให้ผลลัพธ์แตกต่างหรือไม่ บางกรณี อาจต้องมีการศึกษาด้าน cost-effectiveness หรือ cost-benefit ร่วมด้วย

ผลประโยชน์ทับซ้อน

Conflict of interest

ควรพิจารณาว่างานวิจัยได้รับการสนับสนุนเงินทุนจากองค์กรใด หากเป็นหน่วยงานของรัฐหรือองค์กรที่ไม่แสวงหากำไร จะมีความน่าเชื่อถือมากกว่าจากบริษัทผลิตยาหรือเวชภัณฑ์ อย่างไรก็ตาม ปัจจุบันมีกฎเกณฑ์ควบคุมที่ค่อนข้างเข้มงวด และผู้วิจัยต้องประกาศ (declare) ในบทความว่าเคยได้รับการสนับสนุนจากหน่วยงานใดบ้าง ช่วยให้ การพิจารณาทำได้ง่ายขึ้น

สรุป

ทั้งหมดนี้เป็นหลักการและวิธีการเบื้องต้นในการประเมินคุณค่าของผลงานวิจัยว่ามีความน่าเชื่อถือเพียงใด และจะสามารถนำไปประยุกต์ใช้ให้เหมาะสมกับบริบทของเวชปฏิบัติได้หรือไม่ นอกจากนั้น หลายหน่วยงานยังเผยแพร่ checklist ที่ช่วยในการทำ article appraisal สำหรับงานวิจัยประเภทต่างๆ ซึ่งสามารถค้นหาได้ทั่วไป อย่างไรก็ตาม ปัจจุบันมีงานวิจัยที่แตกแขนงไปในรูปแบบใหม่ๆ อีกมาก แพทย์จึงควรศึกษาเพิ่มเติมอยู่เสมอ เพื่อให้เท่าทันความเปลี่ยนแปลงที่รวดเร็วอย่างมากในปัจจุบัน

เอกสารอ้างอิง

1. Al-Jundi A, Sakka S. Critical appraisal of clinical research. *J Clin Diag Res* 2017;11: JE01-05.
2. Schardt C, Adams MB, Owens T, et al. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007; 7:16. doi:10.1186/1472-6947-7-16
3. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005; 95 Suppl 1; S144-50.