

เอกสารประกอบการสอน
หลักการพิจารณางานวิจัยเกี่ยวกับการตรวจวินิจฉัยมาประยุกต์ในเวชปฏิบัติ
(Evidence-based medicine on Diagnostic study)

พญ. อติพร อิงค์สาธิต

“เครื่องมือในการตรวจวินิจฉัย” หรือ “*Diagnostic tools*” หมายถึงสิ่งซึ่งช่วยในการวินิจฉัยโรคหรือภาวะผิดปกติต่างๆ ของร่างกาย ซึ่งอาจเป็นการตรวจทางห้องปฏิบัติการ ข้อมูลทางคลินิกที่ได้จากการซักประวัติ ตรวจร่างกายผู้ป่วย หรือแม้แต่ภาพถ่ายทางรังสีต่างๆ ก็ได้ทั้งสิ้น แพทย์สามารถนำเครื่องมือในการตรวจวินิจฉัยมาใช้ในทางเวชปฏิบัติได้หลายประการ ตัวอย่างเช่นการนำผล renal artery ultrasound มาใช้ในผู้ป่วย chronic kidney disease เพื่อช่วยในการวินิจฉัยภาวะหลอดเลือดไตตีบ (Renal artery stenosis) และยังสามารถนำมาผลมาช่วยบอกความรุนแรงของหลอดเลือดไตตีบ พยากรณ์ในการเกิดหลอดเลือดไตตีบ บอกการตอบสนองต่อการรักษาด้วยการขยายหลอดเลือดด้วยบอลูน ใช้ในการคัดกรองผู้ที่มีภาวะเสี่ยง และใช้ในการติดตามหลังการรักษาว่ามีหลอดเลือดไตตีบอีกหรือไม่[1]

ในเอกสารนี้เน้นการประเมินงานวิจัยเกี่ยวกับการทดสอบเครื่องมือในการตรวจวินิจฉัย ซึ่งมีหลักการพิจารณาดังนี้

1. พิจารณาระเบียบวิธีวิจัยว่าน่าเชื่อถือหรือไม่
2. พิจารณาผลของงานวิจัยว่ามีผลมากน้อยเพียงใด
3. พิจารณาความเป็นไปได้และประโยชน์จากการนำไปใช้ในการวินิจฉัยผู้ป่วยกรณีศึกษาจริง

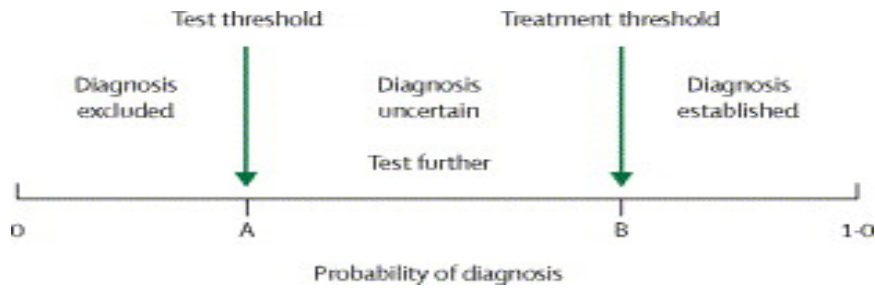
1. การพิจารณาระเบียบวิธีวิจัยว่าน่าเชื่อถือหรือไม่[2]

การประเมินความน่าเชื่อถือของงานวิจัยเกี่ยวกับการตรวจวินิจฉัยมีข้อที่ควรคำนึงถึงในการพิจารณาดังต่อไปนี้

1.1. การศึกษานั้นเลือกกลุ่มตัวอย่างที่นำมาทดสอบเหมาะสมหรือไม่

การตรวจวินิจฉัยควรใช้เมื่อผลการวินิจฉัยที่ได้มีผลต่อการรักษาต่อไป ถ้ากลุ่มตัวอย่างที่ทดสอบมีโอกาสเป็นโรคน้อยมาก (Below A) สามารถตัดโรคนั้นออกไปโดยไม่ต้องทำการตรวจด้วยเครื่องมือทดสอบ หรือถ้ากลุ่มตัวอย่างที่ทดสอบมีโอกาสเป็นโรคค่อนข้างแน่นอนอยู่แล้ว (Above B) ก็ไม่จำเป็นต้องทำการตรวจด้วยเครื่องมือทดสอบเช่นกันและควรจะดำเนินการรักษา แพทย์ควรตรวจผู้ที่มีความผิดปกติอื่นที่มีอาการคล้ายคลึงกัน (A and B) และมีความจำเป็นต้องวินิจฉัยแยกโรคจากกัน ในทางคลินิกเพื่อช่วยแยกแยะผู้ป่วยโรคนั้นๆ จากผู้ที่มีความผิดปกติอื่นจึงจะได้รับประโยชน์จากเครื่องมือทดสอบนั้น ดังรูปที่ 1

รูปที่ 1 Test and treatment thresholds in the diagnosis process[3]



ตัวอย่างเช่น การศึกษาผลของการตรวจวินิจฉัยโดยตรวจ Renal artery ด้วยเครื่อง Doppler ultrasonography ในผู้ป่วยเบาหวานที่มีภาวะไตวายเรื้อรังพบภาวะหลอดเลือดไตตีบ (Renal artery stenosis) ได้บ่อย ผลของงานวิจัยนี้ทำให้มีความหวังว่าการตรวจ Doppler ultrasonography น่าจะมีประโยชน์ในการวินิจฉัยโรคหลอดเลือดไตตีบ แต่เมื่อมีการศึกษาซ้ำโดยใช้กลุ่มตัวอย่างที่ใกล้เคียงกับผู้ป่วยที่นำการตรวจนี้ไปใช้จริงทางคลินิก ได้แก่ผู้ป่วยเบาหวานทั่วไปที่ยังไม่มีภาวะแทรกซ้อน กลับพบว่าประสิทธิภาพของการตรวจนี้ลดลงมาก ทั้งนี้เป็นอิทธิพลของการเลือกใช้กลุ่มตัวอย่างที่มีผลต่อการแปลผลการตรวจวินิจฉัย

1.2 มีการเปรียบเทียบผลของการตรวจวินิจฉัยโดยเครื่องมือที่ต้องการทดสอบกับผลของการตรวจวินิจฉัยด้วยเครื่องมือมาตรฐานโดยการปกปิดผลวินิจฉัย (Blinding) หรือไม่

1.2.1 แพทย์จะต้องค้นคว้า ทบทวนให้ทราบว่าการตรวจวินิจฉัยใดที่ถือเป็นการตรวจวินิจฉัยมาตรฐานที่เหมาะสมต้องสามารถให้การวินิจฉัยได้ถูกต้องตามความเป็นจริงมากที่สุดที่เป็นที่ยอมรับ การในระดับสากลทั่วโลก “Gold Standard” โดยใช้การตรวจที่ดีที่สุดเท่าที่เป็นไปได้และได้รับการยอมรับจากกลุ่มผู้เชี่ยวชาญในสาขาวิชาที่เกี่ยวข้องเป็น Gold Standard ตัวอย่างของการตรวจที่สามารถถือเป็น Gold Standard ได้แก่

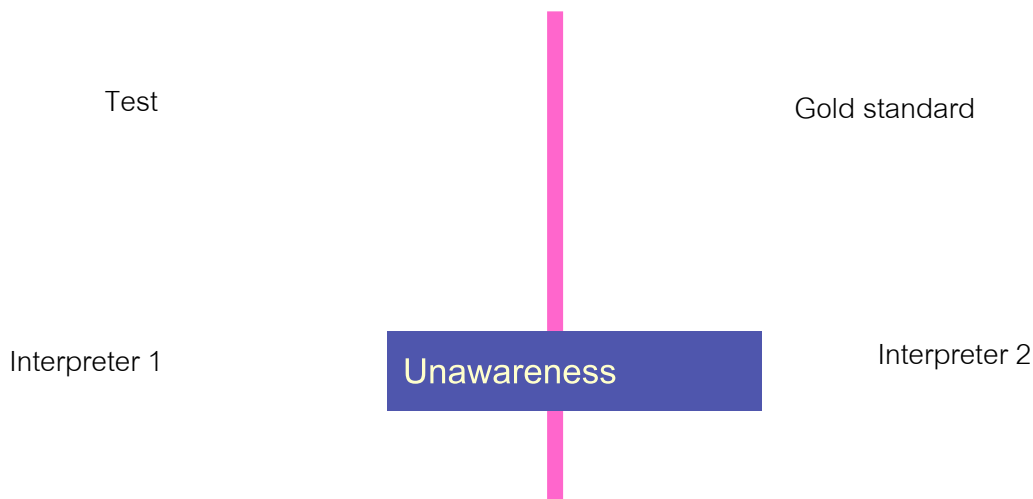
- การตรวจชิ้นเนื้อทางพยาธิวิทยาเป็น Gold Standard สำหรับการวินิจฉัยมะเร็ง
- การเพาะเชื้อเป็น Gold Standard สำหรับการวินิจฉัยโรคติดเชื้อ
- การฉีดสารทึบเข้าหลอดเลือดเป็น Gold Standard สำหรับการวินิจฉัยโรคหลอดเลือดไตตีบ

ดังนั้นผู้อ่านจะต้องพิจารณาและตัดสินใจว่าการตรวจที่ใช้เป็นเกณฑ์อ้างอิงในงานวิจัยนั้นสามารถยอมรับเป็นมาตรฐานที่เหมาะสมได้หรือไม่ หากยอมรับไม่ได้ผลของงานวิจัยนั้นย่อมมีความน่าเชื่อถือ น้อยลง

1.2.2 พิจารณาว่าเครื่องมือที่ต้องการศึกษาต้องไม่เป็นหัวข้อหนึ่งหรือส่วนหนึ่งในการวินิจฉัยมาตรฐาน เช่น ถ้าผู้วิจัยสนใจการตรวจ Antinuclear antibody (ANA) ว่ามีความไวเพียงใดในการวินิจฉัยโรค SLE โดยใช้เกณฑ์ในการวินิจฉัยของ American Rheumatism Association ที่กำหนด 4 ใน 11 ข้อ ซึ่งรวมข้อการตรวจพบ ANA อยู่ในเกณฑ์นี้ด้วย จะทำให้การประเมินความไวของการตรวจ ANA มีค่าสูงกว่าความเป็นจริง

1.2.3 พิจารณาการแปลผลการตรวจของแต่ละวิธี โดยต้องไม่ทราบผลการตรวจของอีกวิธีหนึ่ง ทั้งนี้เพื่อลดอคติในการวัด เพราะถ้าไม่มีการปกปิดการวินิจฉัยผู้อ่านผลย่อมมีแนวโน้มที่จะแปลผลการตรวจใกล้เคียงกับการตรวจด้วยวิธีอ้างอิง ซึ่งจะนำไปสู่การสรุปผลที่ไม่ถูกต้องในภายหลัง ดังนั้นผู้แปลผลทั้งการตรวจที่ต้องการศึกษาและการตรวจวินิจฉัยมาตรฐานจะต้องอิสระซึ่งกันและกันดังรูปที่ 2

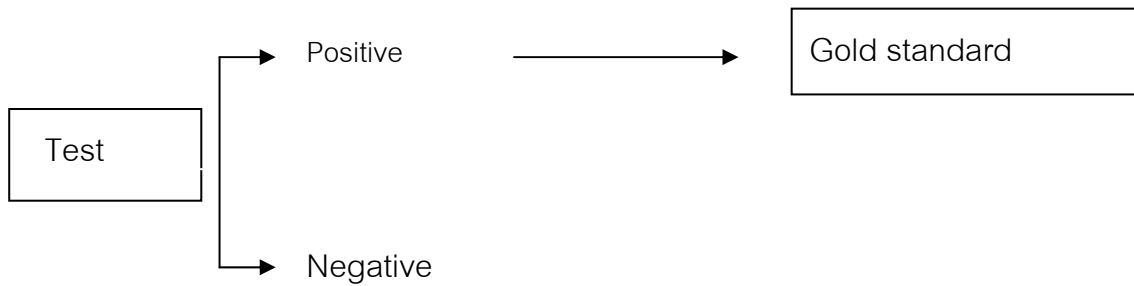
รูปที่ 2 Blinding, independent reading



1.3 ผลการตรวจโดยเครื่องมือที่ต้องการทดสอบมีอิทธิพลต่อการตัดสินใจในการเลือกส่งผู้ป่วยตรวจด้วยวิธีมาตรฐานที่ใช้อ้างอิงในงานวิจัยนั้นหรือไม่

คำถามข้อนี้ต้องการค้นหาสถานการณ์ที่เรียกว่า “Verification bias” หรือ “work – up bias” ซึ่งเกิดขึ้นเมื่อผลของการตรวจที่ต้องการศึกษาวิจัยมีผลต่อการเลือกส่งผู้ป่วยตรวจด้วยวิธีที่ใช้เป็นมาตรฐาน ตัวอย่างเช่นผู้ป่วยที่สงสัยว่าจะเป็นโรคหลอดเลือดไตตีบ ที่มีผลการตรวจ Doppler ultrasonography ให้ผลบวก ย่อมมีโอกาสถูกส่งตรวจ Renal artery angiography ซึ่งเป็นการตรวจมาตรฐานมากกว่าผู้ป่วยที่มีอาการคล้ายคลึงกันแต่ผลการตรวจ exercise test ให้ผลเป็นลบ ทั้งนี้เนื่องจากการตรวจ Renal artery angiography แม้จะเป็นการตรวจมาตรฐานแต่เป็นการตรวจที่มีความยุ่งยาก และมีความเสี่ยงในการเกิดภาวะแทรกซ้อนจากการตรวจได้ ดังนั้นในทางปฏิบัติแพทย์ผู้ดูแลรักษาผู้ป่วยย่อมรู้สึกเล็งเลที่ จะส่งผู้ป่วยที่มีโอกาสเป็นโรคหลอดเลือดไตตีบต่ำไปทำการตรวจ Renal artery angiography ต่อ จึงส่งผลให้ผลการทดสอบประสิทธิภาพของการตรวจวินิจฉัยนั้นบิดเบือนไปได้มาก ดังแสดงแผนภูมิในรูปที่ 3

รูปที่ 3 Verification bias or work-up bias



ดังนั้นในกรณีที่ผู้วิจัยไม่สามารถให้ผู้เข้าร่วมวิจัยมีโอกาสโดยเท่าเทียมกันที่จะได้รับการตรวจโดยวิธีที่เป็นมาตรฐานจนถึงที่สุดให้ได้ทุกคนเนื่องจากข้อจำกัดจากความยุ่งยากและความเสี่ยงของการตรวจมาตรฐาน ผู้วิจัยควรแสดงให้เห็นมาตรการอื่นใดที่จะค้นหาวินิจฉัยโรคซึ่งอาจเป็นการติดตามกลุ่มตัวอย่างที่ไม่ได้รับการตรวจโดยวิธีที่เป็นมาตรฐานด้วยระยะเวลาที่เพียงพอในการที่จะวินิจฉัยโรคโรคนั้นได้เป็นการช่วยทดแทนในการค้นหาโรคที่ต้องการวินิจฉัยให้ถึงที่สุดจึงจะเป็นการศึกษาที่ค้นหาเชื่อถือได้

2. การพิจารณาผลของงานวิจัยว่ามีผลมากน้อยเพียงใด[4]

ก่อนที่จะพิจารณาผลการศึกษาผู้อ่านต้องถามตัวเองทุกครั้งว่าให้น้ำหนักความเชื่อถือในระเบียบวิธีการวิจัยตามหลักพิจารณาเบื้องต้นเท่าไร หากมากกว่าร้อยละ 60 ผู้อ่านจึงค่อยทำการพิจารณาผลของงานวิจัยต่อ แต่หากน้อยกว่านั้นอาจจะต้องขังใจว่าสมควรอ่านต่อหรือไม่เนื่องจากงานวิจัยที่มีปัญหาในด้านความน่าเชื่อถือภายในการศึกษานั้น (Internal validity) ย่อมไม่สามารถนำไปประยุกต์ในเวชปฏิบัติได้

ผลการศึกษาเกี่ยวกับการตรวจวินิจฉัยสามารถจัดวางตัวเลขเข้าในตาราง 2 x 2 ดังรูปที่ 4 ได้ ทั้งนี้เพื่อใช้ในการคำนวณหา ค่าความไว(Sensitivity) ความจำเพาะ(Specificity) ความถูกต้อง(Accuracy), Predictive values และ Likelihood Ratios

รูปที่ 4 Simple properties

		การวินิจฉัยจากการตรวจมาตรฐาน Gold standard	
		เป็นโรค	ไม่เป็น
ผลของการตรวจที่ ต้องการทดสอบ Diagnostic test	ผลบวก	a True positive	b False positive
	ผลลบ	c False negative	d True negative

จากตาราง 2 x 2 จะสามารถคำนวณค่าต่างๆ ได้จากสูตรต่อไปนี้

Sensitivity	=	$a / a + c$
Specificity	=	$d / b + d$
Accuracy	=	$[a + d] / [a + b + c + d]$
Positive Predictive value	=	$a / a + b$
Negative Predictive value	=	$d / c + d$
Likelihood Ratios for positive test	=	$[a / a + c] / [b / b + d]$
Likelihood Ratios for negative test	=	$[c / a + c] / [d / b + d]$
Prevalence or Pre-test probability	=	$[a + c] / [a + b + c + d]$
Pre-test odds	=	Pre-test probability/(1-pre-test probability)
Post-test odds	=	Pre-test odds X Likelihood ratio
Post-test probability	=	Post-test odds/(post-test odds + 1)
Diagnostic odds ratio	=	LR+/LR-

การแปลผลค่า Sensitivity and Specificity

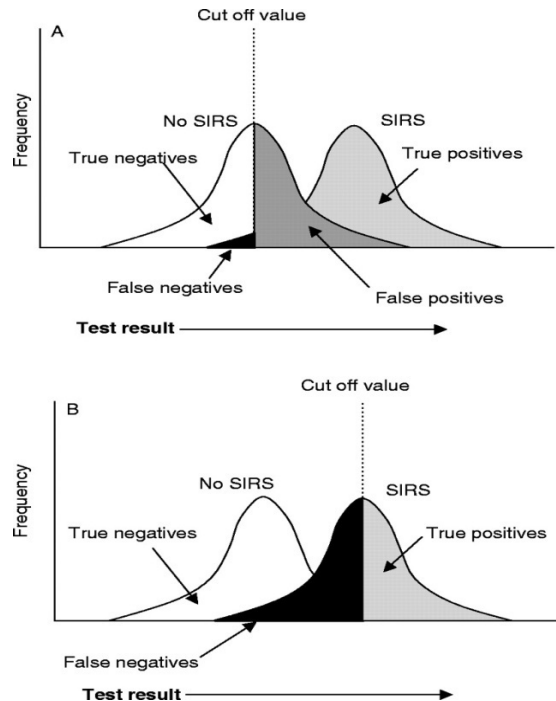
ค่าSensitivity หรือ “true positive rate” ของการตรวจคือสัดส่วนของผู้ป่วยที่ให้ผลการทดสอบเป็นบวกต่อผู้ป่วยทั้งหมด ในทางปฏิบัติควรเลือกใช้การตรวจที่มีค่า Sensitivity สูงๆ ในการตรวจคัดกรองผู้ป่วยสำหรับโรคที่มีความรุนแรงมากแต่สามารถรักษาได้ ซึ่งหากผู้ป่วยไม่ได้รับการวินิจฉัยทั้งที่เป็นโรคจะทำให้ผู้ป่วยเสียประโยชน์มาก นอกจากนี้ยังเหมาะจะใช้เป็นการตรวจเบื้องต้นเพื่อลดจำนวนผู้ป่วยที่จะต้องทำการตรวจที่จำเพาะขึ้นเพื่อการวินิจฉัยต่อไป ผลการตรวจด้วยวิธีการที่มีค่าSensitivity สูงจะมีความหมายมากในกรณีที่ผลการทดสอบเป็นลบเนื่องจากหมายความว่าผู้ป่วยรายนั้นมีโอกาสป่วยด้วยโรคที่วินิจฉัยด้วยการตรวจนั้นน้อยมาก **“SNOUT”**

เครื่องมือในการตรวจวินิจฉัยที่มีค่าSpecificity หรือ “true negative rate” สูง หมายความว่าผู้ป่วยที่ให้ผลการตรวจเป็นบวกมีโอกาสที่จะเป็นโรคจริงสูง ดังนั้นจึงมีประโยชน์ในการยืนยันการวินิจฉัยในกรณีที่มีข้อมูลจากการตรวจอื่นชี้แนะมาบ้างแล้วว่าผู้ป่วยน่าจะป่วยด้วยโรคนั้น การตรวจที่มีคุณสมบัตินี้มีประโยชน์ในกรณีที่ผลบวกกลางจะก่อให้เกิดผลเสียแก่ผู้ป่วยอย่างมากทั้งด้านจิตใจ และการรักษาเสี่ยงต่ออันตราย เช่นกรณีของโรคมะเร็ง หรือทางด้านค่าใช้จ่ายในการรักษา เป็นต้น ดังนั้นการตรวจที่มีSpecificity สูงจึงมีประโยชน์มากในกรณีที่ผลการทดสอบเป็นบวก **“SPIN”**

โดยทั่วไปเรามักจะคาดหวังให้วิธีการตรวจวินิจฉัยที่พัฒนาขึ้นมีทั้ง Sensitivity และ Specificity สูงสุดแต่มักเป็นไปได้ไม่ได้ เมื่อเพิ่ม Sensitivity ให้สูงขึ้น การตรวจนั้นมักมี Specificity ลดลง ในทางตรงกันข้ามการตรวจที่มี Specificity สูงขึ้นมักมี Sensitivity ต่ำ ตัวอย่างเช่น การศึกษาการวัด pro-brain natriuretic peptide (N-BNP) เพื่อวินิจฉัยภาวะ Systemic Inflammatory Response Syndrome (SIRS) โดย Kerbaul และ

ฉะนั้น ถ้าเลือกจุดตัดที่ค่าต่ำดังรูป 5 A จะมีสัดส่วนของผู้ป่วยที่ให้ผลการทดสอบเป็นบวกทั้งที่ไม่มีภาวะนี้ (False positive) มากทำให้เครื่องมือในการตรวจวินิจฉัยนี้มี high sensitivity แต่ low specificity

รูปที่ 5 Test results for two patient populations with different cut off value [5]

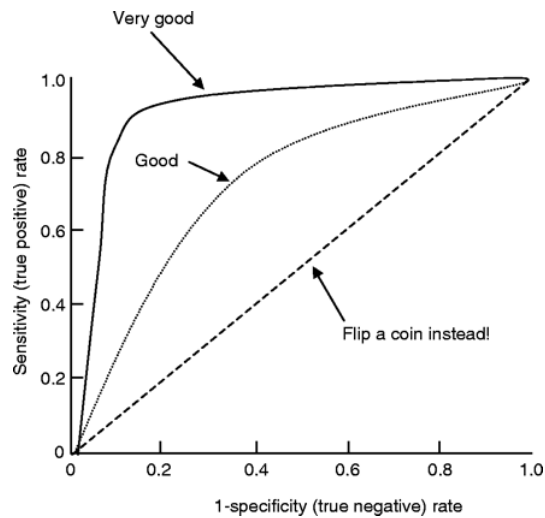


แต่ถ้าเลือกจุดตัดที่มีค่าสูงขึ้นรูป 5 B จะพบสัดส่วนของผู้ป่วยที่ให้ผลการทดสอบเป็นลบทั้งที่เป็นภาวะนี้ (False negative) มากทำให้เครื่องมือในการตรวจวินิจฉัยนี้มี low sensitivity แต่ high specificity ดังนั้นตำแหน่งจุดตัด (cut – off point) ที่เหมาะสม ในการใช้แยกแยะระหว่างภาวะปกติกับความผิดปกติในกรณี que การตรวจให้ผลเป็นข้อมูลต่อเนื่อง (continuous data) ขึ้นอยู่กับความเหมาะสมว่าต้องการให้การตรวจนั้นมีความไวสูงหรือมีความจำเพาะสูง

Receiver Operating Characteristic (ROC) Curves

วิธีการที่สามารถนำมาใช้เพื่อเลือกจุดตัดที่เหมาะสมได้ คือการสร้าง Receiver Operator Characteristic (ROC) curve คือสร้างกราฟความสัมพันธ์ระหว่าง true positive rate (Sensitivity) กับ false positive rate (1-Specificity) โดยการแปรค่าจุดตัด (cut – off point) ที่ใช้ต่างๆ กัน ดังที่เห็นในรูปที่ 6

รูปที่ 6 Receiver Operating Characteristic (ROC) Curve[6]



เครื่องมือในการตรวจวินิจฉัยควรมี Sensitivity สูง และมี Specificity สูง ซึ่งประการหลังจะทำให้มี false positive rate ต่ำ ส่งผลให้ ROC curve เข้าชิดมุมซ้ายบนมากที่สุด นอกจากนี้การสร้าง ROC curve ยังช่วยในการเปรียบเทียบประสิทธิภาพของการตรวจวินิจฉัยได้ด้วย โดยเปรียบเทียบพื้นที่ใต้เส้นโค้งของการตรวจแต่ละชนิด พื้นที่ใต้โค้งที่มากกว่าแสดงถึงประสิทธิภาพที่สูงกว่า

การแปลผลค่า Predictive values

Positive predictive value (PPV) มีความหมายถึงความน่าจะเป็นที่ผู้ป่วยจะเป็นโรคนั้นจริงเมื่อการตรวจให้ผลบวก การตรวจที่มีความจำเพาะสูงมักจะมีค่า PPV สูง

Negative predictive value (NPV) หมายถึงความน่าจะเป็นที่ผู้ป่วยไม่ได้เป็นโรคนั้นเมื่อการตรวจให้ผลเป็นลบ การตรวจที่มีความไวสูงมักจะมีค่า NPV สูง

ในขณะที่ค่า Sensitivity และ Specificity ของการตรวจใดๆ มีค่าคงที่ ค่า predictive value กลับขึ้นอยู่กับความชุกของโรคนั้นในประชากรที่มีการนำการตรวจวินิจฉัยนั้นๆ ไปใช้ด้วย ดังนั้นแม้ว่าในงานวิจัยส่วนใหญ่ มักคำนวณค่า predictive value ไว้ให้ด้วยผู้อ่านพึงระลึกไว้ว่าค่าเหล่านั้นมักนำไปใช้ในเวชปฏิบัติไม่ได้ เนื่องจากความชุกของโรคนั้นในประชากรทั่วไปมักมีความแตกต่างจากกลุ่มประชากรที่นำมาใช้เป็นกลุ่มตัวอย่างในงานวิจัย และนอกจากนั้นความชุกของโรคยังอาจผันแปรตามกลุ่มอายุ เพศ ถิ่นที่อยู่ และปัจจัยอื่นๆ [7]

การแปลผลค่า Likelihood Ratios

Likelihood ratios เป็นวิธีการแสดงประสิทธิภาพของการตรวจวินิจฉัยอีกวิธีหนึ่งนอกเหนือจากค่าความไว ความจำเพาะ และค่า predictive value โดยมีความหมายถึงอัตราส่วนของความน่าจะเป็นของผลการตรวจในผู้ที่เป็นโรคเปรียบเทียบกับผู้ที่ไม่เป็นโรค

เหตุผลที่ทำให้ Likelihood ratios มีความน่าสนใจและควรสนับสนุนให้เป็นที่รู้จักเนื่องจาก

1. สามารถใช้กับการตรวจวินิจฉัยที่มีรายงานผลมากกว่า 2 กลุ่มได้แก่ ระดับความน่าจะเป็นโรคเส้นเลือดหัวใจอุดตัน และการตรวจวินิจฉัยที่รายงานผลเป็นตัวเลขต่อเนื่อง เช่น ระดับ Prostate-specific antigen (PSA) หรือ ระดับความดันโลหิตได้

2. ค่า Likelihood ratios มีความแปรตามความชุกของโรคในกลุ่มตัวอย่างน้อยกว่าค่า

Predictive values

3. สามารถนำไปใช้ในการตัดสินใจวางแผนการรักษาได้ เนื่องจากบ่งบอกโอกาสในการวินิจฉัย (Post-test odds) ได้

ค่า Prevalence อาจเรียกอีกอย่างหนึ่งว่า Pretest probability ค่านี้สามารถนำมาแปลงเป็นค่า Pretest odds ได้ และจากนั้นก็สามารนำไปหาค่า Posttest odds ได้โดยอาศัยสูตรข้างต้น

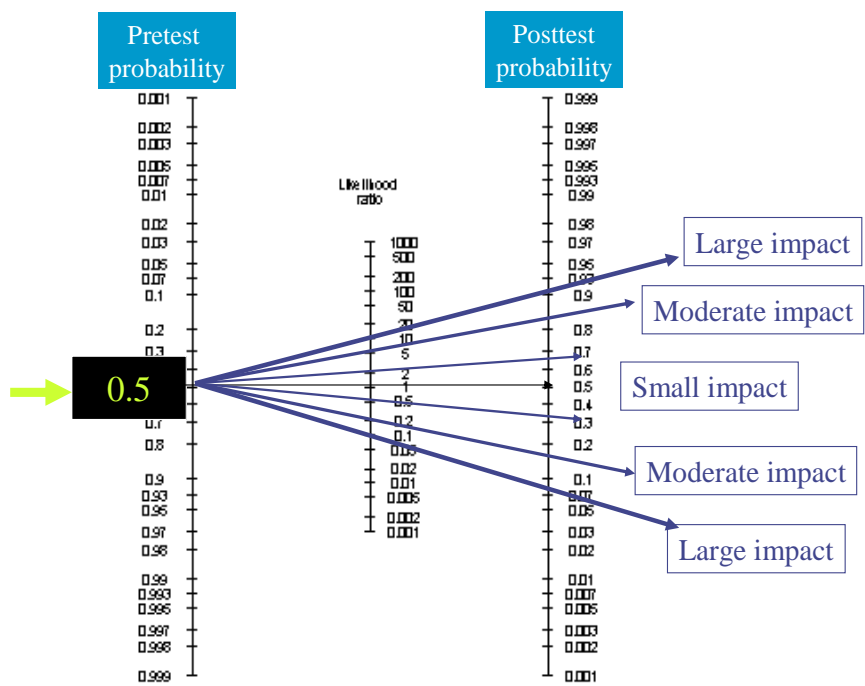
หลังจากนั้นก็สามารนำเอาค่าที่ได้มาแปลงเป็น Posttest probability เพื่อบอกความน่าจะเป็นที่ผู้ป่วยรายนั้นจะเป็นโรคที่ทำการตรวจวินิจฉัยได้

ผู้อ่านต้องทำการค้นหาหรือคำนวณทั้ง Likelihood ratios ในกรณีที่ผลของการตรวจที่ทำการศึกษาเป็นบวก (LR for positive test) และ Likelihood ratios ในกรณีที่ผลของการตรวจที่ทำการศึกษาเป็นลบ (LR for negative test) เพื่อนำมาใช้ในการแปลผลการวิจัยต่อไป

การแปลและการใช้ Fagan nomogram

จะเห็นได้ว่าการคำนวณหาค่า posttest probability ตามวิธีที่แสดงไว้ข้างบนเป็นเรื่องค่อนข้างยุ่งยาก จึงการจัดสร้าง “likelihood ratio nomogram” หรือ “Fagan nomogram” เพื่อทำให้การใช้งานง่ายมากขึ้น

รูปที่ 7 Fagan nomogram[8]



Fagan Nomogram

การใช้งานnomogramนี้ทำได้โดยการใช้ไม้บรรทัดทาบลงไปที่ตัวเลขของPretest probability ที่ต้องการ โดยให้พาดผ่าน Likelihood Ratioของการตรวจวินิจฉัยที่ใช้ด้วยปลายอีกด้านหนึ่งของไม้บรรทัดจะพาดผ่านค่า Posttest probability ให้อ่านค่าได้พอดี

LR for positive test ที่มีค่ามากกว่า 10 หรือต่ำกว่า 0.1 ของ LR for negative test มักทำให้เกิดการเปลี่ยนแปลงของค่า posttest probability ไปจากค่า pretest probability เป็นอย่างมาก ถ้าค่านี้อยู่ระหว่าง 5 - 10 หรือ 0.1 - 0.2 จะทำให้เกิดการเปลี่ยนแปลงปานกลาง ในขณะที่ถ้าอยู่ระหว่าง 2 - 5 หรือ 0.5 - 0.2 การเปลี่ยนแปลงจะเกิดขึ้นเพียงเล็กน้อย

การตรวจที่ให้ค่า LR for positive test ระหว่าง 1 - 2 หรือ 0.5 - 1 สำหรับ LR for negative test มักไม่ค่อยมีประโยชน์ในการวินิจฉัยทั้งนี้เนื่องจากทำให้ค่า Posttest probability แทบไม่แตกต่างจากค่า Pretest probability

โดยทั่วไปแพทย์มักต้องอาศัยประสบการณ์ที่มีในการคาดเดา Pretest probability ในผู้ป่วยแต่ละราย โดยตั้งอยู่บนข้อมูลเกี่ยวกับ เพศ อายุ อาการและการตรวจพบจากการตรวจร่างกาย และข้อมูลอื่นๆ ที่เป็นความรู้พื้นฐานเดิม ทั้งนี้เนื่องจากการเก็บข้อมูลจริงจากฐานประชากรขนาดใหญ่มีอยู่น้อย มีเพียงบางโรค เช่น โรคหลอดเลือดหัวใจที่มีการศึกษาจากฐานข้อมูลขนาดใหญ่ในระดับชาติของสหรัฐอเมริกาให้อ้างอิงได้ อย่างไรก็ตามการนำเอาข้อมูลนั้นมาใช้จะต้องคำนึงถึงความแตกต่างทางด้านเชื้อชาติและวิถีชีวิตจะมีผลทำให้ค่าเหล่านั้นมีความแตกต่างกันหรือไม่ อย่างไรด้วย ช่วงเชื่อมั่น (Confidence interval) การแสดงผลของการตรวจวินิจฉัยก็เป็นเช่นเดียวกับการแสดงผลในงานวิจัยชนิดอื่นๆ ค่าที่ได้คือค่าโดยการประมาณการจาก

การศึกษาในผู้ป่วยเพียงจำนวนหนึ่ง ดังนั้นย่อมมีความแปรปรวนของค่าต่างๆ เกิดขึ้นได้ ยิ่งขนาดของกลุ่มตัวอย่างที่ใช้มีขนาดเล็กเท่าใดอิทธิพลของความแปรปรวนนี้ต่อผลที่ได้ยิ่งมาก ดังนั้นการแสดงผลค่าช่วงเชื่อมั่น 95% (95% confidence intervals) ไว้ในการแสดงผลด้วยจะช่วยให้ผู้อ่านทราบขอบเขตของผลการศึกษามากขึ้นว่ามีความแม่นยำมากน้อยเพียงใด

3. การพิจารณาความเป็นไปได้และประโยชน์จากการนำไปใช้ในการวินิจฉัยผู้ป่วยกรณีศึกษาจริง

สิ่งที่แพทย์ควรคำนึงถึงเมื่อต้องตัดสินใจส่งผู้ป่วยตรวจนอกเหนือจากประสิทธิภาพในการวินิจฉัยโรคของเครื่องมือในการตรวจวินิจฉัยนั้นมีดังต่อไปนี้

3.1 การตรวจนั้นมีการรายงานวิธีการตรวจวินิจฉัยโดยละเอียดจนสามารถทำการตรวจนี้ได้ ในสถาบันของผู้อ่านและมีความเป็นไปได้ที่จะการแปลผลการตรวจแม่นยำหรือไม่

พิจารณาการให้รายละเอียดของการตรวจเครื่องมือในระเบียบวิธีวิจัย โดยควรมีรายละเอียดครอบคลุมสิ่งที่มีความสำคัญทั้งหมด ทั้งการเตรียมผู้ป่วยที่ถูกต้อง เทคนิควิธีการตรวจ ชนิดของน้ำยาเคมีที่ใช้ ชื่อรุ่นของเครื่องมือ การวิเคราะห์และแปลผลการตรวจ เพียงพอที่ผู้อ่านจะสามารถนำการตรวจนั้นไปใช้ได้จริงในเวชปฏิบัติของตนเองได้ ในบางกรณีที่มีปัญหาทางด้านเทคนิคในการตรวจมากจนทำให้การทำการตรวจวินิจฉัยนั้นซ้ำให้เหมือนเดิมได้ยาก เช่น ความแปรปรวนของน้ำยาหรืออุปกรณ์ที่ใช้ในการตรวจ หรือการตรวจนั้นต้องมีการแปลผลโดยบุคลากรที่มีความเชี่ยวชาญมากแต่สถาบันของผู้อ่านไม่มีผู้ที่มีความชำนาญในเรื่องนั้น แพทย์ผู้ทำการรักษาผู้ป่วยน่าจะตัดสินใจเลือกการตรวจวินิจฉัยอื่นมาใช้แทน

3.2 ผลของการศึกษาเกี่ยวกับการตรวจวินิจฉัยนั้นสามารถใช้ได้กับผู้ป่วยในเวชปฏิบัติของผู้อ่านได้หรือไม่

ผู้อ่านควรพิจารณาได้โดยเปรียบเทียบลักษณะของผู้ป่วยของผู้อ่านว่าคล้ายคลึงกับกลุ่มตัวอย่างของการศึกษานั้นหรือไม่โดยดูจาก Inclusion และ exclusion criteria ในการคัดเลือกอาสาสมัครที่รายงานในระเบียบวิธีวิจัย หรือ พิจารณาจากตารางแสดงข้อมูลพื้นฐานของกลุ่มตัวอย่างที่รายงานในส่วนของผลการวิจัย ในกรณีที่ผู้ป่วยของผู้อ่านมีความแตกต่างจากกลุ่มตัวอย่างมากตรวจนั้นย่อมทำให้การนำมาประยุกต์ใช้ผลการตรวจทำได้ลำบากขึ้น

3.3 ผลของการตรวจที่ได้จะเปลี่ยนแปลงแนวทางการรักษาผู้ป่วยในเวชปฏิบัติของผู้อ่านหรือไม่

ดังที่กล่าวมาเบื้องต้นว่าการตรวจวินิจฉัยใดๆ ก็ตามจะมีประโยชน์ในการวินิจฉัยผู้ป่วยก็ต่อเมื่อความน่าจะเป็นที่ผู้ป่วยจะเป็นโรคนั้นๆ อยู่ในระดับกลางๆ เท่านั้น ซึ่งแพทย์ควรตั้ง Treatment threshold ว่าความน่าจะเป็นของโรคเป็นเท่าใดที่จะเริ่มให้การรักษา เช่น ในการรักษาโรคหลอดเลือดไตตีบตั้ง Treatment threshold ของการรักษาไว้ถ้าหากความน่าจะเป็นของโรคอยู่ที่ 0.9 แพทย์จะทำการขยายเส้นเลือด ถ้าการตรวจด้วย Doppler ultrasonography ให้ค่า LR for positive test เท่ากับ 5 หากแพทย์นำไปใช้กับผู้ที่มีโอกาสที่จะเป็นโรคก่อนการทำการตรวจด้วย Doppler ultrasonography ทำให้เพิ่ม Pre-test probability จาก 0.5 เป็น

0.83 (ดู Fagan nomogram รูป 7 ประกอบ) ทำให้การตรวจวินิจฉัยด้วย Doppler ultrasonography ไม่เปลี่ยนแปลงแนวทางการรักษาในผู้ป่วยรายนี้

3.4 พิจารณาประโยชน์ที่ผู้ป่วยจะได้รับจากการตรวจนั้นกับผลเสียที่เกิดขึ้น

เปรียบเทียบผลเป็นประโยชน์ของการตรวจที่ได้ข้อมูลเพิ่มเติมและนำไปสู่การเปลี่ยนแปลงการรักษากับราคา ความคุ้มค่า และอันตรายที่อาจเกิดจากการตรวจทั้งจากเครื่องมือที่สนใจหรือจากการวินิจฉัยมาตรฐาน

และพิจารณาว่าโรคหรือภาวะนั้นมีวิธีการรักษาที่มีประสิทธิภาพ ในบางภาวะที่ไม่มีการรักษาที่มีประสิทธิภาพในขณะนั้นการส่งตรวจ เพียงเพื่อต้องการทราบผลเป็นสิ่งที่ไม่จำเป็นและไม่มีประโยชน์ต่อผู้ป่วย

สรุป ในสถานการณ์ที่แพทย์ต้องให้การรักษาด้วยความรอบคอบเพื่อประโยชน์สูงสุดแก่คนไข้ และเป็นสถานการณ์ที่ความรู้ หลักฐานทางการแพทย์มีจำนวนมากขึ้นแบบก้าวกระโดด แพทย์ต้องมีหลักในการพิจารณา วิเคราะห์ความน่าเชื่อถือและนำผลการศึกษามาใช้อย่างมั่นใจและสามารถบอกเหตุผลที่เลือกใช้การตรวจวินิจฉัยที่เลือกตรวจในคนไข้ได้อย่างมีหลักการและเหมาะสม

เอกสารอ้างอิง

- [1] Gordon Guyatt DS, Brian Haynes. Evaluating Diagnostic Tests. *Clinical Epidemiology: How to Do Clinical Practice Research*. 3 rd ed. Canada: Lippincott William & Wilkins 2006:273-309.
- [2] Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1994 Feb 2;271(5):389-91.
- [3] Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet*. 2005 Apr 23-29;365:1500-5.
- [4] Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994 Mar 2;271(9):703-7.
- [5] Kerbaul F, Giorgi R, Oddoze C, Collart F, Guidon C, Lejeune PJ, et al. High concentrations of N-BNP are related to non-infectious severe SIRS associated with cardiovascular dysfunction occurring after off-pump coronary artery surgery. *Br J Anaesth*. 2004 Nov;93(5):639-44.
- [6] Galley HF. Editorial II: Solid as a ROC. *Br J Anaesth*. 2004 Nov;93(5):623-6.
- [7] Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet*. 2002 Mar 9;359(9309):881-4.
- [8] Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5):257.

Critical appraisal form: Diagnostic study

Citation:

Guide		Comments
I	Are the results valid?	
1.	Did the clinicians face diagnostic uncertainty?	
2.	Was there blind comparison with an independent goal standard?	
3.	Did the results of the test being evaluated influence the decision to perform the gold standard?	
<i>What is your overall evaluation of the paper on a scale of 0 to 10?</i>		
II	What are the results?	
1.	What likelihood ratios are associated with the range of possible test results?	

Guide		Comments
III	How can I apply the results to my patient care?	
1.	Will the reproducibility of the test result and its interpretation be satisfactory in my setting?	
2.	Are the results applicable to my patient?	
3.	Will the results change my management?	
4.	Will patients be better off a result of this test?	

