

บทที่ 6

การทดสอบเพื่อตรวจคัดกรองโรค (Screening Test)

หัวข้อ

1. ความนำ
2. แนวคิดพื้นฐานของการทดสอบเพื่อตรวจคัดกรองโรค
3. ดัชนีที่ใช้ในการประเมินผลการทดสอบ
4. การเลือกใช้ดัชนีที่เหมาะสม
5. กรณีศึกษาตัวอย่างการวินิจฉัยวัณโรคปอด
6. สรุป
7. เอกสารอ้างอิงและแนะนำให้ศึกษาเพิ่มเติม

1. ความนำ

การดูแลรักษาผู้ป่วย แพทย์หรือเจ้าหน้าที่ผู้เกี่ยวข้องจำเป็นต้องวินิจฉัยผู้ป่วยโดยอาศัยประวัติการเจ็บป่วยและการตรวจร่างกาย ซึ่งถือว่ามีความสำคัญมากที่สุด เพื่อให้ได้ข้อมูลที่เพียงพอในการตั้งสมมติฐานการป่วยให้แคบและชัดเจนที่สุด แล้วพิจารณาส่งตรวจเพิ่มเติมทางห้องปฏิบัติการที่เหมาะสมตามสมมติฐานที่ตั้งไว้ การตรวจทางห้องปฏิบัติการจึงมีความสำคัญเช่นกัน เนื่องจากจะเป็นสิ่งที่ช่วยยืนยันได้ว่าเป็นผู้ป่วยเป็นโรคที่แพทย์หรือผู้ให้บริการด้านสุขภาพคาดไว้หรือไม่ การเลือกใช้วิธีการทดสอบหรือการตรวจทางห้องปฏิบัติการวิธีใด เราจำเป็นต้องทราบว่าวิธีการทดสอบนั้น "ดี" หรือ "ไม่ดี" อย่างเป็นวิทยาศาสตร์ เพื่อจะตัดสินใจได้อย่างถูกต้องว่าควรเลือกวิธีการนั้นหรือไม่ และควรเชื่อผลการทดสอบหรือไม่เพียงใด เราจึงจำเป็นต้องมีข้อมูลอย่างเพียงพอในการตัดสินใจนั้น ข้อมูลที่สำคัญนั้นคือดัชนีต่างๆ ที่ใช้ในการประเมินผลการทดสอบนั่นเอง ดัชนีนี้มีหลายตัว การรู้จักและเข้าใจในคุณลักษณะของดัชนีแต่ละตัวจึงมีความสำคัญและจำเป็นมากสำหรับผู้ให้บริการทางการแพทย์และสาธารณสุข เพื่อจะได้นำมาประเมินว่า

เราควรเชื่อถือผลการทดสอบจากวิธีการทดสอบนั้นๆ เพียงใดและควรตัดสินใจเลือกใช้วิธีการทดสอบดังกล่าวหรือไม่ในการวินิจฉัยผู้ป่วย

2. แนวคิดพื้นฐานของการทดสอบเพื่อตรวจคัดกรองโรค

การทดสอบเพื่อวินิจฉัยหรือตรวจคัดกรองโรคหนึ่งๆ อาจมีหลายวิธี การตัดสินใจเลือกใช้วิธีใด เราต้องทราบและเปรียบเทียบคุณสมบัติหรือคุณลักษณะของวิธีการทดสอบเหล่านั้น คุณลักษณะของวิธีการทดสอบดูจากดัชนีต่างๆ ที่ใช้ในการประเมินผลการทดสอบ ดัชนีเหล่านี้ได้มาจากการเปรียบเทียบการทดสอบมาตรฐานหรืออ้างอิง (Gold standard หรือ Reference test) เสมอ (โปรดดูภาพที่ 1 ประกอบ) การทดสอบมาตรฐานมักเป็นการทดสอบที่มีผลการทดสอบที่มีความน่าเชื่อถือมาก ได้แก่ การเพาะเชื้อ แต่อาจมีราคาแพงมากหรืออาจมีความยุ่งยากในการทำ จึงไม่สามารถใช้ได้เป็นการทั่วไปในการวินิจฉัยผู้ป่วย เราจึงต้องอาศัยการทดสอบอื่นที่ราคาถูกลงกว่าหรือทำได้ง่ายและรวดเร็วกว่า แต่จำเป็นต้องเปรียบเทียบกับมาตรฐานนั้นๆ ก่อนว่ามีความน่าเชื่อถือเพียงใด ดัชนีที่ใช้ในการประเมินผลเปรียบก็มีหลายตัว เราจึงจำเป็นต้องทราบว่าควรเลือกใช้ดัชนีตัวใดด้วย นอกจากนี้ การเปรียบเทียบวิธีการทดสอบที่อาจมีอยู่หลายวิธี เราไม่ได้นำวิธีการทดสอบเหล่านั้นมาเปรียบเทียบกับกันตัวต่อตัวโดยตรง แต่เราจะนำการทดสอบแต่ละวิธีมาเปรียบเทียบกับมาตรฐานเสมอ แล้วนำดัชนีต่าง ๆ ที่ได้มาเปรียบเทียบกับว่าดัชนีของการทดสอบวิธีใดน่าเชื่อถือได้มากกว่า เราจึงจะสามารถตัดสินใจเลือกวิธีการทดสอบต่างๆ ได้ เช่น เราต้องการเปรียบเทียบการวินิจฉัยวัณโรคปอดด้วยการตรวจย้อมสีเสมหะแล้วดูด้วยกล้องจุลทรรศน์ หรือเรียกกันว่าการตรวจเสมหะ AFB (Acid-fast bacilli) หรือการตรวจ Smear เปรียบเทียบกับการใช้ CXR (Chest X-ray) เราไม่นำการตรวจ Smear ไปเปรียบเทียบและคำนวณดัชนีการประเมินผลการทดสอบโดยตรงกับการใช้ CXR แต่เราจะใช้การตรวจ Smear เปรียบเทียบกับการเพาะเชื้อเสมหะหาเชื้อวัณโรค (Sputum culture) ซึ่งถือว่าเป็นการทดสอบมาตรฐาน (Gold standard) แล้วคำนวณดัชนีการประเมินผลการทดสอบออกมาได้ชุดหนึ่ง ในทำนองเดียวกันเราก็ทำการเปรียบเทียบการใช้ CXR กับการเพาะเชื้อเสมหะหาเชื้อวัณโรค และคำนวณดัชนีออกมาได้อีกชุดหนึ่ง หลังจากนั้น เราจะนำดัชนีที่ได้ของการตรวจ Smear และดัชนีของการใช้ CXR มาเปรียบเทียบกับว่าดัชนีของวิธีใดดีกว่ากัน เราจึงสามารถบอกได้ว่า การ

ตรวจ Smear หรือการใช้ CXR จะดีกว่ากันในการตรวจวินิจฉัยวัณโรคปอด แพทย์และผู้ให้บริการด้านสุขภาพก็จะสามารถเลือกวิธีที่เหมาะสมกว่าได้

ภาพที่ 1 การทดสอบเพื่อตรวจคัดกรองโรค (Screening Test)

การทดสอบ Screening test	การทดสอบมาตรฐาน/อ้างอิง (Gold standard/Reference test)		รวม (Total)
	คนป่วย	คนไม่ป่วย	
ผลบวก (Positive)	a = ผลบวกจริง (True positive)	b = ผลบวกเทียม (False positive)	รวมผลบวก a+b
ผลลบ (Negative)	c = ผลลบเทียม (False negative)	d = ผลลบจริง (True negative)	รวมผลลบ c+d
รวม	รวมคนป่วย a+c	รวมคนไม่ป่วย b+d	รวมคนที่ทดสอบ a+b+c+d

(ดัดแปลงจาก นายแพทย์ไพบุลย์ โล่ห์สุนทร. 2547. ระบาดวิทยา. หน้า 311.)

3. ดัชนีที่ใช้ในการประเมินผลการทดสอบ

การประเมินผลว่าการทดสอบวิธีใดๆ ดีหรือไม่ดี น่าเชื่อถือหรือไม่น่าเชื่อถือหรือไม่เพียงใด จำเป็นต้องมีดัชนีชี้วัด ซึ่งมีอยู่หลายตัวด้วยกัน (โปรดดูภาพที่ 1, 2 และ 3 ประกอบ) การเลือกใช้ดัชนีตัวใดจึงต้องพิจารณาถึงคุณสมบัติหรือคุณลักษณะของดัชนีตัวนั้นๆ ว่าเป็นอย่างไร โดยทั่วไปเรามักคุ้นเคยกับดัชนี 2 ตัวที่พบบ่อย คือ ความไวของการทดสอบ (Sensitivity) และความจำเพาะของการทดสอบ (Specificity) แต่ยังมีดัชนีอีกหลายตัวที่เราควรรู้จักและนำมาใช้ประโยชน์

ภาพที่ 2 การทดสอบเพื่อตรวจคัดกรองโรค (Screening Test)

ผลบวกจริง หมายถึง ผลการทดสอบเป็นบวกในคนป่วย (a)

ผลบวกเทียม หมายถึง ผลการทดสอบเป็นบวกในคนไม่ป่วย (b)

ผลลบเทียม หมายถึง ผลการทดสอบเป็นลบในคนป่วย (c)

ผลลบจริง หมายถึง ผลการทดสอบเป็นลบในคนไม่ป่วย (d)

ความไวของการทดสอบ หมายถึง ร้อยละของคนป่วยที่ตรวจได้ผลบวกจากการทดสอบ

ความจำเพาะของการทดสอบ หมายถึง ร้อยละของคนไม่ป่วยที่ตรวจได้ผลลบจากการทดสอบ

ค่าพยากรณ์บวก หมายถึง ร้อยละของการทดสอบที่ได้ผลบวกแล้วมีโอกาสเป็นโรค

(ดัดแปลงจาก นายแพทย์ไพบุลย์ โล่ห์สุนทร. 2547. ระบาดวิทยา. หน้า 310-311.)

ภาพที่ 3 ดัชนีที่ใช้ในการประเมินผลการทดสอบ

ความไวของการทดสอบ (Sensitivity) $= [a/(a+c)] \times 100$

ความจำเพาะของการทดสอบ (Specificity) $= [d/(b+d)] \times 100$

ค่าพยากรณ์บวก (Positive predictive value, PV+) $= [a/(a+b)] \times 100$

ค่าพยากรณ์ลบ (Negative predictive value, PV-) $= [d/(c+d)] \times 100$

ความแม่นยำ (Accuracy) $= [(a+d)/(a+b+c+d)] \times 100$

Likelihood ratio $= [a/(a+c)] / [b/(b+d)]$

ดัดแปลงจาก 1. นายแพทย์ไพบุลย์ โล่ห์สุนทร. 2547. ระบาดวิทยา. หน้า 310-312.

2. กิริมย์ กมลรัตนกุล และคณะ. 2542. หลักการทำวิจัยให้สำเร็จ. หน้า 53-54.

หมายเหตุ ความชุก (Prevalence) $= (a+c)/(a+b+c+d)$

4. การเลือกใช้ดัชนีที่เหมาะสม

โดยทั่วไป ดัชนีที่เราคุ้นเคยและได้ยินอยู่บ่อยๆ คือ **ความไวของการทดสอบ (Sensitivity)** และ **ความจำเพาะของการทดสอบ (Specificity)** หากเราเปรียบเทียบกันระหว่างการทดสอบวิธีต่างๆ วิธีใดที่มีความไวและความจำเพาะสูงกว่า ก็ถือได้ว่ามีความถูกต้องมากกว่าและน่าใช้มากกว่า แต่บางครั้งพบว่าความไวและความจำเพาะของวิธีการทดสอบเหล่านั้น ไม่ได้แตกต่างกันมากนัก หรือไม่แตกต่างกันเลย ทำให้เราตัดสินใจได้ยากกว่าวิธีการทดสอบใดดีกว่า ดัชนีอีกตัวหนึ่งที่มีผู้นิยมใช้กันมาก คือ ค่าพยากรณ์ โดยเฉพาะ **ค่าพยากรณ์บวก (Positive predictive value หรือ PV+)** หากค่าพยากรณ์บวกของวิธีใดมากกว่า ถือว่าดีกว่า อย่างไรก็ตาม **ค่าพยากรณ์บวกนี้ ก็มีข้อจำกัด คือ มีความผันแปรไปตามความชุกของโรค** ดังสถานการณ์สมมติในตัวอย่างภาพที่ 4 และ 5 จะเห็นได้ว่า ในกรณีที่ใช้วิธีการทดสอบเดียวกันที่มีความไวร้อยละ 98 และความจำเพาะร้อยละ 99 แต่ในสถานการณ์จำลองที่มีความชุกของโรคไม่เท่ากัน จะทำให้ค่าพยากรณ์บวกมีค่าไม่เท่ากัน กล่าวคือ ในสถานการณ์ A ที่มีความชุกของโรคร้อยละ 1 ค่าพยากรณ์บวกคำนวณได้ร้อยละ 49.7 แต่ในสถานการณ์ B ที่มีความชุกของโรคร้อยละ 10 สามารถคำนวณค่าพยากรณ์บวกได้ร้อยละ 91.6 นั่นคือ หากมีความชุกของโรคเพิ่มขึ้น จะทำให้ค่าพยากรณ์บวกมีค่าเพิ่มขึ้น อีกตัวอย่างข้อมูลในภาพที่ 6 ก็จะได้เห็นได้อย่างชัดเจนว่า ค่าพยากรณ์บวกมีความผันแปรไปตามความชุกของโรค เมื่อความชุกของโรคเพิ่มมากขึ้น ค่าพยากรณ์บวกก็มีค่าสูงขึ้น แม้ว่าค่าความไวและความจำเพาะมีค่าคงที่ เราจึงสามารถสรุปได้ว่า **ค่าพยากรณ์บวกอาจไม่ใช่ดัชนีที่ดีนักในการประเมินผลการทดสอบวิธีต่าง ๆ เพราะมีความผันแปรไปได้ตามความชุกของโรค**

ภาพที่ 4 ตัวอย่างสถานการณ์ A

< Situation A >

Prevalence rate= 1% in population of 100,000

Sensitivity= 98%, Specificity= 99%

Predictive value of positivity= ?

		Disease		
Test		Yes	No	Total
	Positive	980	990	1970
	Negative	20	98,010	98030
	Total	1000	99,000	100,000

$$\text{P.V (pos.)} = a/(a+b) \times 100 = 980/1970 = 49.7 \%$$

Source: Korean Institute of Tuberculosis (KIT). 2002. Slides from the Training Course on Tuberculosis Control.

ภาพที่ 5 ตัวอย่างสถานการณ์ B

< Situation B >

Prevalence rate= 10% in population of 100,000

Sensitivity= 98%, Specificity= 99%

Predictive value of positivity= ?

		Disease		
Test		Yes	No	Total
	Positive	9,800	900	10,700
	Negative	200	89,100	89,300
	Total	10,000	90,000	100,000

$$\text{P.V (pos.)} = a/(a+b) \times 100 = 9800/10700 = 91.6 \%$$

Source: Korean Institute of Tuberculosis (KIT). 2002. Slides from the Training Course on Tuberculosis Control.

ภาพที่ 6 ค่าพยากรณ์บวก (PV+) และค่าพยากรณ์ลบ (PV-) ตามระดับความชุกของโรคต่าง ๆ เมื่อความไวและความจำเพาะของการทดสอบเท่ากับร้อยละ 90

ความชุกของโรค (ร้อยละ)	ค่าการพยากรณ์บวก	ค่าการพยากรณ์ลบ
1	16.1	99.9
2	27.9	99.9
5	50.0	99.7
10	67.9	99.4
20	82.6	98.7
50	95.0	95.0
75	98.3	86.4
100	100.0	-

(ดัดแปลงจาก นพ.ไพบุลย์ โล่ห์สุนทร. 2547. ระบาดวิทยา. หน้า 319.)

อย่างไรก็ตาม มีค่าดัชนีที่ดีอีกค่าหนึ่งที่ไม่มีความผันแปรหรือไม่เปลี่ยนแปลงไปตามความชุกของโรค คือ **Likelihood ratio** (ภาพที่ 7) ซึ่งเป็นค่าที่บ่งบอกถึงความผิดปกติ (ผลบวก) จากการตรวจพบในกลุ่มคนที่เป็นโรคว่าเป็นกี่เท่าของกลุ่มคนที่ไม่เป็นโรค หากเราลองคำนวณค่าดัชนี Likelihood ratio ในสถานการณ์สมมติในภาพที่ 4 และ 5 จะพบว่า มีค่า Likelihood ratio เท่ากันคือ 98 เท่า แม้ว่าทั้งสองสถานการณ์นั้นมีความชุกของโรคต่างกัน ดังนั้น **Likelihood ratio** จึงควรเป็นดัชนีที่ดีค่าหนึ่ง เพราะไม่เปลี่ยนแปลงไปตามความชุกของโรค หากวิธีการทดสอบใดมีค่า Likelihood ratio ดีกว่า ควรเป็นวิธีการทดสอบที่ดีกว่า

ภาพที่ 7 Likelihood Ratio

$$\text{Likelihood ratio} = [a/(a+c)] / [b/(b+d)]$$

- เป็นค่าที่บ่งบอกถึง ความผิดปกติ (ผลบวก) จากการตรวจพบในกลุ่มคนที่ เป็นโรค เป็น กี่เท่า ของกลุ่มที่ ไม่เป็นโรค
- ไม่ขึ้นกับความชุก (Prevalence) ในกลุ่มตัวอย่างงานวิจัยหรือในชุมชน

(ดัดแปลงจาก ภิรมย์ กมลรัตนกุล และคณะ. 2542. หลักการทํารววจัยให้สำเร็จ. หน้า 54-55.)

5. กรณีศึกษาตัวอย่างการวินิจฉัยวัณโรคปอด

การวินิจฉัยวัณโรคปอดเป็นกรณีตัวอย่างที่ดีกรณีหนึ่ง เนื่องจากองค์การอนามัยโลก หรือ WHO (World Health Organization) แนะนำให้ใช้การตรวจ Smear (การตรวจย้อมสีเสมหะและดูด้วยกล้องจุลทรรศน์) เป็นวิธีการวินิจฉัยผู้ป่วยวัณโรคปอด โดย WHO (2003: 19) ได้กำหนดให้ การตรวจ Smear เป็นองค์ประกอบหนึ่งของกลยุทธ์ DOTS (Directly Observed Treatment, Short-course) ที่ใช้ในการควบคุมวัณโรค แต่แพทย์ส่วนใหญ่มีความคุ้นเคยกับการใช้ CXR (Chest X-ray) มากกว่า จึงเป็นประเด็นถกเถียงกันว่า ควรใช้การตรวจ Smear หรือ CXR ในการวินิจฉัยวัณโรคปอดดีกว่า

เราลองคำนวณดัชนีต่างๆ ที่ใช้ในการประเมินผลการทดสอบ โดยใช้ข้อมูลที่ได้จากการทดลองจริงในประเทศอินเดียเมื่อกว่า 30 ปีมาแล้วในภาพที่ 8 และ 9 เปรียบเทียบการตรวจ Smear (AFB) และ CXR กับการเพาะเชื้อวัณโรค (Culture) ซึ่งถือเป็นวิธีการทดสอบมาตรฐาน/อ้างอิง เราจะได้ค่าดัชนีที่คำนวณได้ในภาพที่ 10

ภาพที่ 8***Correlation of the yield of radiographic examination and sputum culture in patients with clinical signs suggestive of tuberculosis***

Radiography	Culture		
	Positive	Negative	Total
Positive	142	85	227
Negative	20	1982	2002
Total	162	2067	2229

Source: World Health Organization. 2004. *Toman's Tuberculosis Case Detection, Treatment, and Monitoring - Questions and Answers*. Second Edition. Printed in China. WHO/HTM/TB/2004.334. Page 62.

ภาพที่ 9***Correlation of the yields of two cultures and of direct smear microscopy of a single spot specimen in patients with clinical signs suggestive of tuberculosis***

Smear	Culture		
	Positive	Negative	Total
Positive	130	15	145
Negative	32	2052	2084
Total	162	2067	2229

Source: World Health Organization. 2004. *Toman's Tuberculosis Case Detection, Treatment, and Monitoring - Questions and Answers*. Second Edition. Printed in China. WHO/HTM/TB/2004.334. Page 63.

ภาพที่ 10 ดัชนีเปรียบเทียบการวินิจฉัยวัณโรค

	X-ray	Smear
ความไว (Sensitivity, %)	87.7	80.2
ความจำเพาะ (Specificity, %)	95.9	99.3
ค่าพยากรณ์บวก (PV+, %)	62.6	89.7
ค่าพยากรณ์ลบ (PV-, %)	99.0	98.5
ความแม่นยำ (Accuracy)	95.3	97.9
Likelihood ratio (เท่า)	21.3	110.6

หมายเหตุ: คำนวณจากข้อมูลในภาพที่ 8 และ 9

เมื่อเปรียบเทียบดัชนีที่คำนวณได้เปรียบเทียบระหว่าง X-ray (CXR) กับ Smear (AFB) จะพบว่า CXR มีความไวดีกว่า Smear ขณะที่ความจำเพาะของ Smear ดีกว่า CXR เล็กน้อยหรือกล่าวได้ว่าไม่แตกต่างกันชัดเจน ค่าพยากรณ์บวกของ Smear ดีกว่า CXR มาก คือ ค่าพยากรณ์บวกของ Smear มีค่าร้อยละ 89.7 ส่วนของ CXR มีค่าเพียงร้อยละ 62.6 อาจจะแปลได้ว่าการใช้ Smear ในการวินิจฉัยดีกว่า CXR อย่างไรก็ตาม เราทราบแล้วว่า ค่าพยากรณ์บวกมีความผันแปรไปตามความชุกของโรคได้ แต่ค่า Likelihood ratio ไม่เปลี่ยนแปลงไปตามความชุกของโรค เราพบว่า Smear มีค่า Likelihood ratio สูงถึง 110.6 เท่า ขณะที่ CXR มีค่าเพียง 21.3 เท่า นั่นคือ การวินิจฉัยวัณโรคปอดโดยใช้ Smear จึงควรดีกว่าการใช้ CXR ส่วนค่าพยากรณ์ลบและความแม่นยำของทั้ง Smear และ CXR ไม่มีค่าแตกต่างกันชัดเจน

6. สรุป

เมื่อมีวิธีการทดสอบหลายๆ วิธีในการวินิจฉัยโรค เราจำเป็นต้องมีดัชนีในการประเมินผลว่าวิธีการทดสอบใดดีหรือควรใช้มากกว่า ดัชนีที่ใช้มีหลายค่า ได้แก่ ความไว ความจำเพาะ ค่าพยากรณ์บวก Likelihood ratio เป็นต้น ค่าพยากรณ์บวกมีความผันแปรไปได้ตามความชุกของโรค หากความชุกของโรคเพิ่มขึ้น ค่าพยากรณ์บวกก็เพิ่มขึ้น แต่ **Likelihood ratio** ไม่มีการเปลี่ยนแปลงไปตามความชุกของโรค จึงควรเป็นดัชนีในการประเมินผลการทดสอบได้ดีค่าหนึ่ง.

7. เอกสารอ้างอิงและแนะนำให้ศึกษาเพิ่มเติม

1. ไพบุญย์ โล่ห์สุนทร, ศาสตราจารย์ นายแพทย์. 2547. ระบาดวิทยา. พิมพ์ครั้งที่ 5: มกราคม 2547. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
2. ภิรมย์ กมลรัตนกุล, มนต์ชัย ซาลาประวรรตน์, ทวีสิน ต้นประยูร (บรรณาธิการ). 2543. หลักการทำวิจัยให้สำเร็จ. พิมพ์ครั้งที่ 2. บริษัท เท็กซ์ แอนด์ เจอร์นัล พับลิเคชั่น จำกัด.
3. Korean Institute of Tuberculosis. 2002. **Slide Presentation Sets of the Training Course on Tuberculosis Control, 2002.**
4. World Health Organization (WHO). 2004. **K. Toman's Tuberculosis Case Detection, Treatment, and Monitoring: Questions and Answers.** Second Edition. Printed in China . WHO/HTM/TB/2004.334.
5. World Health Organization (WHO). 2003. **Treatment of Tuberculosis: Guidelines for National Programmes.** Third Edition. Italy: Jotto Associati s.a.s.-Biella-Italy. WHO/CDS /TB/2003.313.