

วิธีทดสอบด้านห้องปฏิบัติการเพื่อวินิจฉัยโรค : ความรู้พื้นฐาน การแปลผล และการประเมินคุณค่า

อมร ถิ่นรัมย์ พ.บ.

ปัจจุบันการวินิจฉัยโรคติดเชื้อหลายๆ โรคต้องอาศัยผลการตรวจทางห้องปฏิบัติการเพื่อนำข้อมูลมาวิเคราะห์และวินิจฉัยโรค แม้แต่ในอนาคตก็จะต้องพึ่งผลการตรวจทางห้องปฏิบัติการมากขึ้น ผลการตรวจจากห้องปฏิบัติการมักจะแสดงออกในรูปผลบวกหรือผลลบ ส่วนในกรณีที่วัดผลการตรวจออกมาเป็นตัวเลขวก็มักจะสรุปผลว่าบวกหรือลบเช่นกัน ผลบวกหรือลบดังกล่าวมีอิทธิพลต่อการตัดสินใจของแพทย์ว่าผู้ป่วยเป็นโรคนั้นหรือไม่ โดยทั่วไปถ้าได้ผลบวกก็จะตัดสินใจว่าเป็นโรคนั้น ทำนองเดียวกันถ้าได้ผลลบก็จะตัดสินใจว่าไม่เป็นโรคนั้น ส่วนผู้ที่มีประสบการณ์หรือมีความรู้ในโรคติดเชื้อต่างๆ ก็จะระมัดระวังในการแปลผลบวกและลบมากขึ้น เพราะผลบวกที่ได้ อาจจะเป็นผลบวกเท็จ กล่าวคือ ผู้ป่วยไม่ได้เป็นโรคนั้นแต่ให้ผลบวกกับการตรวจ ในทำนองเดียวกันผลลบที่ได้ อาจจะเป็นผลลบเท็จ กล่าวคือ ผู้ป่วยยังมีโอกาสเป็นโรคนั้นถึงแม้การตรวจตัวอย่างจะให้ผลลบกับการตรวจ ปัญหาในการแปลผลการตรวจเหล่านี้เป็นเรื่องที่ประสบบ่อยทุกวัน หากไม่ทราบถึงคุณสมบัติของวิธีทดสอบแต่ละวิธีแล้วจะทำให้แปลผลการตรวจได้ไม่ชัดเจนหรือผิดพลาดไปจนกลายเป็นข่าวโด่งดังในหน้าหนังสือพิมพ์ได้ นอกจากนี้ความรู้พื้นฐานที่ได้จากบทความนี้จะนำไปประยุกต์ใช้ในการแปลผลการทดสอบอื่นๆ ได้ด้วย บทความนี้จะกล่าวถึงรายละเอียดต่างๆ ของคุณสมบัติของวิธีทดสอบต่างๆ เพื่อให้รู้จักใช้และแปลผลการทดสอบได้ถูกต้องตามลำดับ ดังนี้

- บทนำ (ทำไมจึงต้องมี การทดสอบและ gold standard คืออะไร?)
- ขั้นตอนการทำวิจัยในด้านวิธีทดสอบเพื่อวินิจฉัยโรค
- กลยุทธ์ที่ใช้วินิจฉัยโรคทางคลินิก
- ลักษณะของโรคที่ควรนำวิธีทดสอบมาใช้และความชุกของโรคในสถานการณ์ต่างๆ
- คุณสมบัติของวิธีทดสอบเพื่อวินิจฉัยโรคและการแปลผลการทดสอบ

หน่วยระบาดวิทยาคลินิก ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล

- ตัวอย่างแสดงการคำนวณค่าต่างๆ ที่เป็นคุณสมบัติของวิธีทดสอบเพื่อวินิจฉัยโรค
 - วิธีสร้าง Receiver Operating Characteristics (ROC) curve
 - การใช้จุดตัดหลายจุดและการคำนวณค่า likelihood ratio
- อนึ่ง บทความนี้ไม่ได้มีจุดมุ่งหมายที่จะอธิบายรายละเอียดการทำวิธีทดสอบใดๆ ทั้งสิ้น

บทนำ

การวินิจฉัยโรคเป็นสิ่งจำเป็นอย่างยิ่งในการวางแผนการรักษาผู้ป่วย ความยากง่ายในการวินิจฉัยแต่ละโรคแตกต่างกัน โรคบางชนิดอาศัยข้อมูลจากการซักประวัติและการตรวจร่างกายที่สามารถวินิจฉัยได้อย่างถูกต้อง หรือให้การรักษาที่เหมาะสมต่อไปได้ โรคอีกหลายชนิดยังต้องอาศัยการตรวจค้นเพิ่มเติม. วิธีการตรวจค้นมีทั้งชนิดทำได้ข้างเตียงและในห้องปฏิบัติการซึ่งต้องอาศัยเครื่องมือต่างๆ เมื่อนำข้อมูลที่ได้จากการตรวจค้นเพิ่มเติมมารวมกับข้อมูลที่ได้จากการซักประวัติและการตรวจร่างกายจะสามารถวินิจฉัยโรคได้ในที่สุด. โดยทั่วไปแต่ละโรคจะมีวิธีการวินิจฉัยที่เป็นมาตรฐานอยู่วิธีหนึ่งซึ่งเป็นวิธีที่ใช้วินิจฉัยโรคได้แน่นอน (ถูกต้อง) ที่สุดในปัจจุบัน หรือเป็นวิธีที่ยอมรับทั่วไปสำหรับโรคแต่ละชนิดดังที่เขียนไว้ในตำรามาตรฐาน วิธีการวินิจฉัยหรือวิธีการทดสอบนี้จะเป็ นวิธีที่แพทย์ผู้รักษาทุกคนอยากนำมาใช้โดยเร็วที่สุด เพื่อจะได้ทราบผลการวินิจฉัยที่แน่นอนและใช้ผลการทดสอบมาวางแผนการรักษา เช่น การวินิจฉัยภาวะปอดอักเสบต้องอาศัยอาการไข้ ไอ หอบ และเจ็บป็นในเนื้อปอดที่เห็นได้จากภาพถ่ายรังสีทรวงอกและการตรวจชิ้นเนื้อปอดว่ามี การอักเสบ ในกรณีนี้การซักประวัติ, การตรวจร่างกาย และภาพถ่ายรังสีทรวงอกจะเป็นวิธีมาตรฐานซึ่งมีชื่อว่า gold standard (ตามตำราระบาดวิทยาคลินิก) ในการวินิจฉัยปอดอักเสบเฉียบพลันจะทำให้แพทย์ผู้รักษาวางแผนการรักษานิด

ระดับประคองได้ เช่น การให้ผู้ป่วยพักผ่อน การให้ออกซิเจนถ้าหายใจไม่เพียงพอ แต่ถ้าจะวินิจฉัยไปถึงเชื้อก่อเหตุของปอดอักเสบเฉียบพลัน จะต้องรวมผลการเพาะเชื้อและอื่นๆ ไว้ใน gold standard ด้วย การทราบผลการตรวจหาเชื้อก่อเหตุจะทำให้แพทย์วางแผนการรักษาที่จำเพาะได้ คือ การให้ยาต้านจุลชีพที่เหมาะสม ถ้าเป็นการวินิจฉัยมะเร็งหลอดลมต้องอาศัยการตรวจชิ้นเนื้อจากรอยโรคว่ามีลักษณะเข้าได้กับเซลล์มะเร็ง วิธีการตรวจชิ้นเนื้อจึงเป็นวิธีวินิจฉัยมะเร็งได้แน่นอนที่สุดและจัดเป็น gold standard เช่นกัน. การทำ CPC ระหว่างภาควิชาด้านคลินิกกับภาควิชาพยาธิวิทยาจึงเป็นการวัด clinical standard กับ gold standard นั้นเอง ถึงแม้วิธีการวินิจฉัยโรคที่เป็น gold standard จะเป็นวิธีที่แพทย์อยากนำมาใช้กับผู้ป่วยมากที่สุด แต่ในทางปฏิบัติปรากฏว่าวิธีวินิจฉัยที่เป็น gold standard ส่วนมากจะเป็นวิธีที่มีปัญหาในด้านอื่นๆ ดังนี้

1.1 ใช้เวลานานหลายวันกว่าจะทราบผล

1.2 มีความเสี่ยงต่อการเกิดความพิการมากน้อยตลอดจนอาจถึงแก่กรรมได้ทั้งในผู้ป่วยหรือผู้ป่วย เมื่อนำวิธีการนี้มาใช้

1.3 มีความยุ่งยากซับซ้อนหรือเกิดความไม่สะดวกในการใช้วิธีดังกล่าว

1.4 มีราคาแพง หรือต้องใช้เครื่องมือราคาแพง

ถ้าวิธีวินิจฉัยที่เป็น gold standard สำหรับโรคชนิดหนึ่งไม่มีปัญหาดังกล่าวแล้ว วิธีนั้นจะเป็นวิธีทดสอบที่นำมาใช้ทั่วไปและไม่มีความจำเป็นที่จะต้องคิดค้นหาวิธีวินิจฉัยใหม่ เช่น การตรวจเสมียร์เลือดหาเชื้อมาลาเรีย แต่ถ้าวิธีวินิจฉัยที่เป็น gold standard ยังประสบปัญหาดังกล่าวก็จะมีผู้คิดค้นหาวิธีใหม่ที่สะดวก รวดเร็ว ง่าย ราคาถูก โดยที่ความถูกต้องอาจจะดีกว่า, เท่ากัน หรือดีกว่าวิธีมาตรฐาน จึงเกิดการประเมินคุณค่าของวิธีทดสอบใหม่ที่คิดค้นขึ้น โดยเปรียบเทียบกับวิธีทดสอบมาตรฐาน (gold standard) ดังจะได้กล่าวต่อไป

2. ปัจจัยต่างๆ ที่เกี่ยวข้องกับการทำวิจัยในด้านวิธีทดสอบเพื่อวินิจฉัยโรค

การวิจัยทั่วไปต้องอาศัยปัจจัยหลายอย่าง ที่สำคัญคือ ต้องมีบุคลากรหลายระดับที่มีความสามารถในการช่วยปฏิบัติงานด้านต่างๆ โดยเฉพาะเมื่อทำงานวิจัยขนาดใหญ่ งานด้านต่างๆ ที่กล่าวถึงมีตั้งแต่ผู้ให้คำปรึกษา ผู้เชี่ยวชาญในด้านที่งานวิจัยภาคพื้นถึงผู้ช่วยปฏิบัติงานวิจัย ผู้เก็บข้อมูลและรวบรวม เจ้าหน้าที่พิมพ์ดีดและเย็บเล่ม เป็นต้น ปัจจัยที่สำคัญมากอีกอย่างหนึ่งคือ ทุนทรัพย์ในการปฏิบัติงาน ซึ่งต้องมีมากพอที่จะผลักดันให้งานดำเนินไปด้วยดี ส่วนตัวผู้ทำวิจัยจะต้องมีทั้งเวลาและความตั้งใจที่แน่วแน่ในการทำงานวิจัยให้สำเร็จ

2.1 คำถามวิจัย เมื่อตัดสินใจจะทำวิจัยด้านวิธีทดสอบเพื่อวินิจฉัยโรค จะต้องทบทวนเอกสารในเรื่องนั้นให้ต้องมาก่อน เพื่อ

ให้เกิดความคิดในการวางแผนการศึกษาและทำให้การศึกษามีสมบูรณ์ โดยที่ไม่เกิดความผิดพลาดดังที่เคยปรากฏในงานวิจัยที่ดีพิมพ์แล้ว หลังจากทบทวนในเรื่องที่สนใจแล้ว จะต้องพิจารณาดูให้รอบคอบอีกครั้งว่าผลงานวิจัยจะเป็นประโยชน์มากเพียงพอที่จะนำมาวิจัยหรือไม่ และเป้าหมายหลักของงานวิจัยคืออะไรซึ่งเป็นคำถามวิจัยของงาน คำถามที่ดีจะระบุเป้าหมายของงานวิจัยได้อย่างชัดเจน คำถามวิจัยสำหรับการทดสอบเพื่อวินิจฉัยโรคจะอยู่ในแนวทางดังนี้ว่า "วิธีทดสอบ ก จะช่วยวินิจฉัยโรค ก ได้ดีเพียงใด" คำถามวิจัยที่มีเนื้อความละเอียดมากขึ้นจะช่วยให้ผู้ทบทวนงานวิจัยที่มีประสบการณ์เห็นภาพพจน์แผนงานวิจัยได้ดี และบางครั้งช่วยให้นักสถิติคำนวณขนาดของประชากรศึกษาที่เหมาะสมได้ การนํ่วงคยหรือถกปัญหาในเรื่องคำถามวิจัยกับผู้เชี่ยวชาญ อาจเกิดประโยชน์มากเพราะทำให้ผู้วิจัยมองเห็นภาพพจน์ของงานวิจัยชัดเจนขึ้น นอกจากนี้ยังเป็นการทบทวนความคิดของผู้วิจัยในหัวข้อและขอบเขตของงานวิจัยให้ชัดเจนขึ้น สารที่จะขยายในคำถามวิจัยดังกล่าวข้างต้นคือ "วิธีทดสอบ ก" คือวิธีอะไร? จะต้องใช้วิธีทดสอบ ก ในโรค ก ที่มีความรุนแรงเพียงใด หรือในระยะใดของโรค ก ส่วนข้อความที่ว่า "ดีเพียงใด" คืออะไรและหมายถึงอะไร? เพราะคำว่า "ดีเพียงใด" อาจจะมีหมายถึง

1. วินิจฉัยโรคได้แม่นยำถึงร้อยละ 85 หรือ

2. คุ่มค่ากับเงินที่ลงทุนในการจัดทำวิธีทดสอบ หรือ

3. เกิดประโยชน์แก่ผู้ป่วยมากที่สุด

ซึ่งความหมายแต่ละอย่างจะกำหนดแนวทางวิจัยที่แตกต่างกัน คำถามวิจัยที่ดีจะดึงดูดความสนใจของผู้ร่วมงานวิจัยและของผู้ที่พิจารณาให้ทุนงานวิจัยด้วย ส่วนการเขียนสมมติฐานงานวิจัยจะคล้ายคลึงกับคำถามงานวิจัย เพียงแต่ว่าการเขียนสมมติฐานจะขยายเนื้อหาที่เป็นคำตอบของคำถามวิจัยที่ถามไว้ ซึ่งเป็นการเสนอแนวทางการวิจัยในการแก้ปัญหาที่ตั้งเป็นคำถามวิจัยนั้น

2.2 ลักษณะและขนาดประชากรเป้าหมาย ประชากรเป้าหมายหรืออ้างอิง คือ ประชากรทุกคนซึ่งมีลักษณะที่เข้ากับหลักเกณฑ์ที่วางไว้และเป็นประชากรกลุ่มที่มีปัญหาในการวินิจฉัยจนต้องนำวิธีทดสอบมาใช้ตรวจตัวอย่าง ถ้าประชากรดังกล่าวมีเป็นจำนวนมากทั่วโลกนี้ เราอาจจะกำหนดลักษณะประชากรให้แคบลง เช่น กำหนดลักษณะของโรคให้ตรงกับปัญหาที่เกิดขึ้นบ่อยที่สุดหรือสำคัญมากที่สุด เช่น ถ้าต้องการศึกษาในประชากร (ผู้ป่วย) ที่ติดเชื้ออหิวาตกโรค ก็กำหนดลักษณะของประชากรเป้าหมายว่าเป็นผู้ป่วยที่ไอแห้งๆ และมีไข้มานานเกิน 7 วัน เพราะผู้ป่วยกลุ่มนี้เป็นปัญหาในการวินิจฉัยโรค เป็นต้น อย่างไรก็ตามแม้จะได้กำหนดลักษณะผู้ป่วยไว้แล้ว บางครั้งเราก็ยังไม่สามารถรวบรวมผู้ป่วยทุกรายมารับการศึกษาได้ จึงต้องเลือกผู้ป่วยบางรายมาศึกษา ซึ่งต้องอาศัยวิธีสุ่มเลือกที่เหมาะสมเพื่อให้ได้ตัวแทน (ประชากรตัวอย่าง) ที่ถูกต้องของประชากรกลุ่มนี้ (ประชากรเป้าหมาย) การได้ตัวแทนที่ถูกต้องมาศึกษาจะทำให้เราสามารถนำผลการศึกษากลับไป

ประยุกต์ใช้กับผู้ป่วยทั่วไปได้อย่างไม่คลาดเคลื่อน การนำวิธีทดสอบไปตรวจตัวอย่างในประชากรทั่วไปจึงต้องมีวิธีสุ่มตัวอย่างประชากรที่เหมาะสม เท่าที่ประสบมาพบว่าการศึกษาทางคลินิกโดยทั่วไปมักจะมีจำนวนผู้ป่วยน้อยจนไม่ต้องสุ่มเลือก หรือมีน้อยเกินไปจนต้องศึกษาผู้ป่วยจากแหล่งต่างๆ (multi-centre study) ถ้าวิธีทดสอบนี้ใช้เป็น screening test ในประชาชน ควรพิจารณาว่าโรคที่ต้องการตรวจจะพบได้ในประชากรที่มีลักษณะอย่างไร? เช่น มีอายุช่วงใดแล้วให้ค่าจำกัดความของประชากรที่จะนำมาศึกษาตามช่วงอายุ ถ้านำวิธีทดสอบมาใช้กับโรคควรพิจารณาว่าสถานภาพใดของโรคนี้ที่เป็นปัญหาจนต้องหาวิธีทดสอบมาตรวจค้น เพราะถ้าผู้ป่วยมีอาการแสดงของโรคชัดเจนก็ไม่ต้องใช้การทดสอบใดๆ อีก หรือเป็นโรคที่ยังไม่แสดงอาการทางคลินิก แต่การตรวจพบก่อนจะทำให้การรักษาได้ก่อนที่จะเกิดความเสียหาย โดยทั่วไปค่าจำกัดความของผู้ป่วยมักจะเป็นในรูปแบบที่กำหนดให้ตรวจพบโรคได้ตั้งแต่ระยะแรกเริ่ม หรือในระยะที่โรคแสดงออกไม่ชัดเจน หรือระยะที่ผู้ป่วยมีอาการเหมือนกับโรคอื่นๆ อีกหลายโรคจนแยกกันไม่ออก ค่าจำกัดความนี้ครอบคลุมไปถึงประชากรกลุ่มควบคุมด้วย (control group) เช่น คณะผู้วิจัยศึกษาผู้ป่วยที่มีอายุระหว่าง 13-40 ปี และติดเชื้อมัยโคพลาสมาในระบบทางเดินอากาศหายใจส่วนล่างที่แสดงอาการไอแห้งๆ อย่างน้อย 7 วัน และมีไข้ ประชากรที่นำมาศึกษาทั้งกลุ่ม "มีโรค" และ "ไม่มีโรค" จะต้องมีอาการไข้และไอแห้งๆ อย่างน้อย 7 วัน และมีอายุอยู่ระหว่าง 13-40 ปีเหมือนกัน ค่าจำกัดความดังกล่าวจะเป็น inclusion criteria สำหรับนำผู้ป่วยเข้ามศึกษาในการประเมินคุณค่าของวิธีการทดสอบ

ขนาดของกลุ่ม (จำนวน) ประชากรศึกษาทั้งกลุ่มที่ "มีโรค" และ "ไม่มีโรค" จะได้จากสูตรโดยคำนวณแยกกลุ่มกัน ทั้งนี้เพราะ sensitivity และ specificity รวมทั้งค่าคาดหวังสูงสุดและต่ำสุด (R และ L) มักจะมีค่าแตกต่างกัน เมื่อคำนวณได้จำนวนประชากรในแต่ละกลุ่มก็นำมารวมกันเป็นจำนวนประชากรที่ควรนำมารับการทดสอบทั้งหมด การคำนวณแยกกลุ่มกันนี้จะใช้ในกรณีที่น่าวิธีการวินิจฉัยมาเป็น diagnostic test

สูตรที่ใช้คำนวณขนาดประชากรในแต่ละกลุ่มได้แก่

$$n = \frac{(15.4) P (1 - P)}{(R - L)^2}$$

ถ้าคำนวณจำนวนกลุ่ม "มีโรค"

ค่าต่างๆ และคำอธิบายมีดังนี้

n = จำนวนประชากรในกลุ่ม "มีโรค"

R และ L เป็นค่าคาดหวังที่สูงสุดและต่ำสุด เมื่อมีความมั่นใจว่าการศึกษาหาค่า sensitivity ในประชากรกลุ่มนี้จะให้ค่าถูกต้องถึงร้อยละ 95 ในกลุ่มมีโรค กล่าวคือ

R คือค่าคาดหวังสูงสุดของ sensitivity สำหรับการทดสอบวิธีนี้คิดเป็น probability เมื่อมีความมั่นใจดังกล่าว เช่น คาดว่าจะมีค่าร้อยละ 90, R จะเท่ากับ 0.9

L คือค่าคาดหวังต่ำสุดของ sensitivity สำหรับการทดสอบวิธีนี้คิดเป็น probability เมื่อมีความมั่นใจดังกล่าว เช่น คาดว่าจะมีค่าร้อยละ 70, L จะเท่ากับ 0.7

P = ค่าเฉลี่ยที่คาดหวังของ sensitivity สำหรับการทดสอบวิธีนี้คิดเป็น probability เช่น คาดว่าค่าเฉลี่ยอยู่ที่ร้อยละ 80, P จะเท่ากับ 0.8 และค่า P จะต้องอยู่ระหว่างค่า R และ L เสมอ จากนั้นให้นำค่าต่างๆ ลงไปแทนในสูตร จะได้ค่า n = 62 ราย (ไม่เชื่อก็คงคำนวณดู) นั่นคือ ประชากรกลุ่ม "มีโรค" ที่จะนำมาศึกษาควรมีจำนวน 62 ราย

ส่วนการคำนวณจำนวนประชากรศึกษาสำหรับกลุ่ม "ไม่มีโรค" จะมีขั้นตอนการคำนวณแบบเดียวกัน เพียงแต่เปลี่ยนค่า sensitivity เป็น specificity และค่าคาดหวังของ R และ L ตามค่า specificity ที่คาดหวัง

ในที่สุดจำนวนประชากรที่ควรใช้ในการศึกษาครั้งนี้จะเป็นผลรวมของจำนวนประชากรของทั้ง 2 กลุ่มนั่นเอง

จำนวนประชากรศึกษาทั้งกลุ่ม "มีโรค" และ "ไม่มีโรค" ที่น้อยเกินไปจะทำให้ผลการศึกษาค่า sensitivity และ specificity มีโอกาสคลาดเคลื่อนจากความเป็นจริงมากขึ้น ถ้าใช้จำนวนประชากรมากเกินไปจะมีการสูญเสียของแรงงาน เวลา และทรัพย์สินส่วนหนึ่ง เพราะถ้าใช้จำนวนประชากรที่พอเหมาะจะได้ค่าถูกต้องตรงหรือใกล้เคียงความเป็นจริงอยู่แล้ว

ถ้ามีการทดสอบไปตรวจหาประชากรที่เป็นโรคในชุมชน การคำนวณหาจำนวนประชากรที่เหมาะสมในชุมชนอาจจะทำโดยคำนวณหาจำนวนประชากรที่เหมาะสมในกลุ่มมีโรคก่อน จากนั้นให้คาดการณ์ว่าความชุกของโรคนั้นควรมีค่าเท่าใดในชุมชน จากนั้นให้นำจำนวนประชากรกลุ่มมีโรคเป็นตัวตั้งและหารด้วยค่าความชุกของโรคนั้นในชุมชน เช่น ความชุกของโรคนั้นในชุมชนที่ต้องการศึกษามีร้อยละ 1 และจากการคำนวณประชากรกลุ่มมีโรคสมมติว่าต้องใช้จำนวน 62 ราย ดังนั้นจำนวนประชากรในชุมชนที่ต้องนำมาศึกษาจะมีจำนวนทั้งสิ้น $= \frac{62 \times 100}{1} = 6,200$ คน

วิธีคำนวณจำนวนประชากรแบบนี้ จึงใช้กับการทดสอบที่นำมาใช้แบบ screening test ซึ่งนำไปใช้ในชุมชนนั่นเอง

2.3 ความแปรปรวนของข้อมูล (ค่าตัวแปร) เครื่องมือที่นำมาใช้วัดค่าต่างๆ ในการทดสอบจะต้องตรวจสอบความเหมาะสมในการนำมาวัดเพื่อตอบคำถามวิจัย เครื่องมือที่มีราคาแพงมากจะมีผู้ใช้ไม่แพร่หลายเท่าเครื่องมือขนาดเล็ก น้ำหนักเบา และราคาถูก ถ้ามีเครื่องมือหลายอย่างที่วัดได้ เครื่องมือที่มี coefficient of variation ต่ำสุดจะเป็นเครื่องมือที่เหมาะสมที่สุด วิธีวัดผลต้องสะดวก ทำง่าย ได้ผลเร็ว ถ้ายังเป็นวิธีใหม่จะต้องซักซ้อมและบรรยายขั้นตอนการใช้และการวัดให้ชัดเจน การควบคุมคุณภาพในการวัดมีความสำคัญมาก การใช้ตัวอย่างมาตรฐานทั้งชนิดที่ให้ผลบวกและผลลบ หรือที่มีน้ำหนักมาตรฐานมาร่วมทดสอบด้วยจะป้องกันอคติจากการวัด (measurement bias) การวัดตัวอย่าง

เดียวกันซ้ำๆ หลายๆ ครั้งจะต้องได้ค่าตรงกัน (repeatable, reproducible) และถูกต้องกับความจริง (valid) ถ้าเครื่องมือสามารถวัดค่าตัวแปรออกเป็นตัวเลขจะตัดปัญหาในด้าน diagnostic suspicion bias ลงได้ เพราะเครื่องมือเป็นตัวอ่านผลลัพธ์ ในกรณีที่มีการผ่านผลการทดสอบยังขึ้นอยู่กับการสังเกตและความนึกคิดของคนอ่าน และคนอ่านเป็นผู้ตัดสินใจให้ผลการทดสอบเป็นผลบวกหรือลบ หรือให้คะแนนเป็น 1+ ถึง 4+ เป็นต้น ผู้ที่อ่านผลจะต้องอ่านผลการทดสอบโดยไม่ทราบที่มาของตัวอย่าง เพื่อหลีกเลี่ยงอคติ (diagnostic suspicion bias) ที่จะอ่านได้เข้ากับผู้ "มีโรค" หรือ "ไม่มีโรค" การวัดค่าตัวแปรใดๆ ที่ไม่ได้ผลลัพธ์ออกมาเป็นค่าแบบต่อเนื่อง (continuous) จะต้องตั้งหลักเกณฑ์การอ่านผลทดสอบทั้งบวกและลบให้ชัดเจน และแสดงให้เห็นว่าไม่มีปัญหาในด้านของ observer variation ในการอ่านผลการทดสอบ โดยให้ผู้อื่นมาร่วมอ่านผลด้วยและต้องอ่านได้ผลตรงกัน.

ความแปรปรวนของข้อมูล (ค่าตัวแปร, variable) ที่เก็บแบ่งออกได้เป็น 2 ประเภท คือ

2.3.1 ความแปรปรวนตามธรรมชาติ (Biological variation) ซึ่งจัดเป็น random variation เช่น การวัดแรงดันเลือด การวัดระดับน้ำตาล หรือการวัดระดับ cortisol ในคนเดียวกันแต่วันเวลาต่างกัน ก็ได้ค่าแตกต่างกัน การแก้ไขที่จุดนี้ทำได้โดยกำหนดเวลาเก็บตัวอย่างให้ตรงกัน หรือจัดให้ผู้ถูกวัดอยู่ในสภาพเดียวกัน หรือสภาพที่สากหลอมรับ เพื่อให้ค่าที่วัดได้ระหว่างกลุ่มต่างๆ เปรียบเทียบกันได้ เมื่อจัดสภาพและกำหนดเวลาให้เหมือนกันแล้วควรเก็บและวัดตัวอย่างในคนเดียวกัน 2-3 ครั้งและหาค่าเฉลี่ยต่อไป

2.3.2 ความแปรปรวนอย่างเป็นระบบ (Systematic variation) จัดเป็นความแปรปรวนที่ผู้วิจัยต้องพยายามกำจัด เพราะเป็นความผิดพลาดทางเทคนิค บางคนเรียกความผิดพลาดนี้เป็น systematic bias ซึ่งเป็นผลร้ายต่อการศึกษา ความแปรปรวนชนิดนี้เป็นอคติที่ป้องกันได้และต้องไม่ให้เกิดขึ้นในงานวิจัยเพราะจะทำให้ค่าที่วัดได้ผิดไปจากความจริงทั้งหมดทั้งๆ ที่ดูอย่างผิวเผินก็เห็นว่าเป็นข้อมูลที่น่าเชื่อถือได้ ทั้งนี้เพราะการตรวจซ้ำยังได้ค่าใกล้เคียงกัน (reproducible) ตัวอย่างของ systematic bias ซึ่งในกรณีของการวัดนี้เรียกว่า measurement bias ได้แก่ เครื่องวัดแรงดันเลือดที่เสื่อมสภาพ วัดแรงดันเลือดทุกครั้งก็ได้ค่าต่ำกว่าความเป็นจริง แต่ค่าที่วัดได้ก็ยังใกล้เคียงกันถ้าวัดในคนเดียวกัน หรือน้ำยาที่เสื่อมคุณภาพซึ่งทำให้วัดได้ค่าหรือไตเตอร์ต่ำกว่าที่เป็นจริงทุกครั้ง. แม้แต่การปนเปื้อนของ pyrogen ในน้ำยาทำให้ผลการทดสอบบางอย่างได้ค่าสูงขึ้นหรือต่ำลง สิ่งต่างๆ เหล่านี้ล้วนทำให้เกิดอคติชนิดนี้ได้ การแก้ไข systematic bias ในการทดสอบคือการนำตัวอย่างที่ทราบค่าแน่นอนและมีความเข้มข้นต่างๆ กัน มาทดสอบร่วมกับตัวอย่างที่เก็บได้ซึ่งเป็นการควบคุมคุณภาพของการทดสอบนั่นเอง (quality control) ถ้ามีการทดสอบหลายขั้นตอนจะต้องพิจารณาวิธีการควบคุมคุณภาพในแต่ละขั้นตอนด้วย ฟิงระลึกไว้เสมอ

ว่าการตรวจได้ผลลบไม่ได้แปลว่าตัวอย่างที่นำมาตรวจไม่มีสารที่ต้องการตรวจเท่านั้น แต่ยังหมายความว่าเครื่องมือหรือน้ำยาที่ใช้ อาจจะไม่สามารถตรวจหาสารนั้นที่มีอยู่ในตัวอย่างได้ ซึ่งอาจจะเกิดจากการที่มีจำนวนสารนั้นน้อยมากจนตรวจไม่พบ หรือมีจำนวนพอเหมาะ แต่น้ำยาและเครื่องมือเสื่อมคุณภาพ หรือมีสารนั้นมากเกินไปจนกลับให้ผลลบ เช่น ในกรณีของ Prozone phenomenon เป็นต้น การวัดผลใดๆ ที่ต้องอาศัยสมรรถภาพของบุคลากรจะต้องตรวจสอบว่า บุคลากรเหล่านั้นมีความสามารถในการอ่านผลตามหลักเกณฑ์ที่วางไว้ โดยกำหนดหลักเกณฑ์และนำตัวอย่างที่ให้ผลต่างๆ กันมาให้อ่านทดสอบหรือให้ผู้อ่านเปรียบเทียบกับตัวอย่างให้เห็นชัดเจน มิฉะนั้นอาจจะเกิด systematic bias ได้ เช่น ผลสุดท้ายของปฏิกิริยาชนิดหนึ่งที่ทำให้ผลบวก จะเปลี่ยนจากการที่สารละลายไม่มีสีมาเป็นสีเหลือง แต่เราอาศัยคนตาบอดสีเหลืองมาอ่านผลก็จะอ่านผิดเป็นผลลบ เป็นต้น การอ่านผลที่ยุ่งยากกว่านี้มีอีกมาก เช่น การอ่านผลการเป็นวันของน้ำยา การอ่านความขุ่นหรือการจับตัวของเม็ดเลือดแดงในการทดสอบการอ่านผลเหล่านี้ จะต้องกำหนดเกณฑ์ให้ชัดเจน.

จุดตัดของวิธีทดสอบใหม่ที่แบ่งค่าตัวแปรออกเป็นผลบวกหรือลบ จะต้องใช้หลักเกณฑ์ที่เหมาะสมเพื่อให้เป็นจุดตัดที่เกิดประโยชน์ต่อผู้เข้ารับการทดสอบมากที่สุด หรือให้ตรงเป้าหมายการวิจัยมากที่สุด ค่าปกติที่อ้างถึงอาจจะได้มาจากการศึกษานำร่องหรือจากตำรา เพื่อให้ผู้ทำนายไปใช้หรือแปลผลการทดสอบทราบที่มาของค่าปกติและนำไปประยุกต์ใช้ได้อย่างถูกต้อง

เมื่อได้กำหนดหลักเกณฑ์ต่างๆ เรียบร้อยแล้ว ผู้วิจัยควรทำแบบฟอร์มสำหรับเก็บรวบรวมข้อมูลต่อไป โดยเฉพาะแบบฟอร์มที่อำนวยความสะดวกทั้งในการบันทึกข้อมูลและป้อนเข้าเครื่องคอมพิวเตอร์ได้เลย จะทำให้ประหยัดเวลาในการวิเคราะห์ข้อมูลอย่างยิ่ง

2.4 สถิติที่ใช้ในการวิเคราะห์ข้อมูลในงานวิจัยด้านการวินิจฉัยโรค สถิติจะเริ่มจากการแสดงข้อมูลพื้นฐานของกลุ่มประชากรที่นำมาศึกษาทุกกลุ่มทั้งกลุ่ม "มีโรค" และ "ไม่มีโรค" ข้อมูลพื้นฐานดังกล่าวได้แก่ เพศ อายุ โรคเดิมที่มีอยู่ ฐานะครอบครัว ระดับการศึกษา นิสัยส่วนตัว การเปรียบเทียบในด้านข้อมูลพื้นฐานจะทำให้ผู้อ่านเข้าใจสภาพของประชากรศึกษาที่รับเข้ามาทำการศึกษา และเป็นการแสดงให้เห็นว่าข้อมูลพื้นฐานของประชากรศึกษาแต่ละกลุ่มไม่แตกต่างกัน ข้อมูลพื้นฐานที่นำมาแสดงจะเป็นข้อมูลที่มีผลต่อการทดสอบหรือต่อผลลัพธ์ด้วย โดยทั่วไปใช้สถิติประเภท T-test, ANOVA, Chi-square ยืนยันความคล้ายคลึงกันของข้อมูลพื้นฐานของประชากรแต่ละกลุ่ม ส่วนการตรวจหาคุณสมบัติประจำการทดสอบแต่ละวิธีส่วนใหญ่จะใช้ตาราง 2 x 2 (two by two table) เป็นหลักพื้นฐาน และใช้ข้อมูลในตาราง 2 x 2 เพื่อแปลผลการทดสอบด้วย จากตาราง 2 x 2 จะสามารถคำนวณ sensitivity, specificity, positive predictive value (อ่านในหัวข้อที่ 5) ฯลฯ

ถ้ามีตัวแปรหลายชนิดที่ใช้ในการวินิจฉัยโรคเดียว จะต้องใช้คอมพิวเตอร์ช่วยวิเคราะห์ เช่น ใช้โปรแกรม multivariate analysis, logistic regression เพื่อคัดเลือกตัวแปรที่มีประโยชน์มากที่สุด และรองลงมาตามลำดับในการวินิจฉัยและตัดตัวแปรที่ไม่มีประโยชน์ออก ผู้วิจัยควรพิจารณารายละเอียดในด้านนี้กับนักสถิติตั้งแต่เริ่มต้นวางแผนงานวิจัย ถ้าต้องการวิเคราะห์ถึงระดับนโยบายการตัดสินใจจะรับหรือไม่รับการทดสอบวิธีนั้นมาใช้ในสถานพยาบาลแห่งหนึ่ง จะต้องอาศัยหลักวิเคราะห์ทางด้าน Decision Analysis ร่วมด้วย

เมื่อพิจารณาถึงขั้นตอนนี้แล้ว ผู้ทำงานวิจัยจะสามารถเขียนโครงการงานวิจัย (research proposal) ได้ การเขียนโครงการงานวิจัยจะเป็นการทบทวนขั้นต้นและแนวทางดำเนินงานวิจัยให้เห็นแจ่มชัดหรือเข้าใจดีขึ้น ระหว่างการเขียนอาจจะทำให้ผู้วิจัยพบเห็นปัญหาบางอย่างที่ซ่อนเร้นอยู่ซึ่งจะได้หยิบยกปัญหาดังกล่าวมาแก้ไขก่อน จึงเท่ากับเป็นการเตรียมพร้อมก่อนลงมือปฏิบัติการจริง บางครั้งอาจจะต้องทดลองทำวิจัยนำร่องไปก่อน เพื่อให้เนื้อหาบางจุดในงานวิจัยเกิดผลที่เห็นเด่นชัดขึ้น ทำให้สามารถให้รายละเอียดได้ดีในโครงการงานวิจัยก่อนจะนำมาเสนอต่อที่ประชุม ผู้ที่ใช้เวลาเตรียมการอย่างดี ชักซ้อมแนวทางปฏิบัติ และการประสานงานจนเกิดความเข้าใจเหมือนกับที่จะปฏิบัติจริงในงานวิจัยขั้นนั้น จะเท่ากับลงมือปฏิบัติงานวิจัยสำเร็จไปแล้วครึ่งหนึ่ง โครงการงานวิจัยที่สมบูรณ์จะมีแม้กระทั่งตารางต่างๆ ที่เตรียมช่องว่างไว้บันทึกหรือแสดงข้อมูลที่ต้องการเก็บ และเตรียมการอภิปราย และกิตติกรรมประกาศไว้พร้อมกับเอกสารอ้างอิงด้วย รายละเอียดต่างๆ ในบทเสนองานวิจัยและการจัดรูปแบบบทความแต่ละตอนได้อย่างสวยงาม จะทำให้การพิจารณาอนุมัติงานวิจัยง่ายขึ้น ปัจจุบันนิยมพิมพ์เก็บโดยใช้เครื่องไมโครคอมพิวเตอร์และ word processor ไทย-อังกฤษ หรือ Wordstar ซึ่งสะดวกทั้งในการแก้ไขคำผิดและการจัดรูปแบบบทความในการเขียนโครงการงานวิจัย

3. กลยุทธ์ที่ใช้วินิจฉัยโรคทางคลินิก

การวินิจฉัยโรคผู้ป่วยรายใด แพทย์ต้องอาศัยประวัติการเจ็บป่วยและผลการตรวจร่างกายผู้ป่วยอย่างละเอียดลอมารประกอบเป็นข้อมูลพื้นฐานในการวินิจฉัยโรค บางครั้งเมื่อมาถึงขั้นการตรวจร่างกายก็ตรวจพบลักษณะเฉพาะบางอย่างที่ใช้วินิจฉัยโรคได้ทันที เช่น ผู้ป่วยดับม้ามโต และมีลักษณะใบหน้าเหมือนผู้ป่วยโรคเบตาฮีธาลัสซีเมีย, ลักษณะหน้าตาของผู้ป่วยที่เป็น Down's syndrome รอยโรคที่ถุงอัมพาตผู้ป่วยรายหนึ่งที่มีไข้เรื้อรังมา 2 สัปดาห์ (เป็น scrub typhus) เป็นต้น วิธีการวินิจฉัยโรคเช่นนี้เรียกว่า pattern recognition ในกรณีนี้แพทย์อาจจะงดใช้การทดสอบทางห้องปฏิบัติการเพื่อวินิจฉัยโรคเพราะสามารถวางแผนการรักษาได้แล้วที่จริงแล้วประวัติหรือการตรวจร่างกายที่พบลักษณะเฉพาะสำหรับโรคบางอย่าง เป็นวิธีทดสอบเพื่อวินิจฉัยโรคเช่นกัน และเป็นวิธีที่ประหยัดที่สุด มีคุณค่ามากเหลือคณานับ แต่ใช้เวลาและความละเอียด

ลอของผู้ตรวจ แพทย์ที่มีประสบการณ์มากจะได้เปรียบกว่าผู้ต้องประสบการณ์และใช้ประโยชน์ข้อนี้ได้เต็มที่

เมื่อข้อมูลจากการซักประวัติและการตรวจร่างกายไม่สามารถบ่งชี้โรคใดโรคหนึ่ง การทดสอบทางห้องปฏิบัติการเพื่อวินิจฉัยโรคจะให้ข้อมูลเพิ่มเติมในการวินิจฉัยต่อไป กลยุทธ์ในตอนนี้สามารถใช้ได้ 3 วิธี วิธีแรกคือ ดำเนินการตรวจคัดกรอง algorithm ที่มีผู้วางไว้เรียกว่า arborization strategy วิธีที่สองคือ ค้นคว้าเพิ่มเติมจากตำราต่างๆ และ "หว่าน" การทดสอบไปทั่ว ซึ่งเป็นวิธีที่เสียเวลาและสิ้นเปลืองมาก เรียกว่า exhaustion method ส่วนวิธีสุดท้ายที่นำใช้คือ hypotheticodeductive strategy ซึ่งเป็นวิธีที่แพทย์นึกถึงโรคต่างๆ ที่เป็นได้จากอาการสำคัญและผลการตรวจร่างกายแล้วหาข้อมูลจากการทดสอบเพื่อใช้สนับสนุนหรือแยกโรคออกไป วิธีทดสอบเพื่อวินิจฉัยโรคมีอยู่มากมายตั้งแต่วิธีที่ทราบผลเร็ว ราคาถูกไป จนถึงวิธีที่ทราบผลช้าและราคาแพง วิธีที่ทราบผลเร็วให้ผลแม่นยำและราคาถูกมักจะเป็นที่นิยมและนำมาใช้ก่อนวิธีการทดสอบที่ยุ่งยากและทราบผลช้า ปัจจุบันนี้เทคโนโลยีต่างๆ เจริญรุดหน้าจนมีปัญหาเรื่องค่าใช้จ่ายในการทดสอบเพราะมีราคาแพง ถึงแม้จะให้ผลการทดสอบที่แม่นยำเชื่อถือได้ก็เป็นปัญหาในการให้ผู้ป่วยได้รับการทดสอบชนิดนั้น

การประเมินคุณค่าและรู้จักแปลผลการทดสอบเพื่อวินิจฉัยโรคจึงมีความจำเป็นเพื่อให้เกิดประโยชน์เต็มที่แก่ผู้ป่วย และได้ผลตรงจุดมุ่งหมายของผู้ส่งตรวจ

4. ลักษณะของโรคที่ควรนำวิธีทดสอบมาใช้และความชุกของโรคตามสถานการณ์ต่างๆ

วัตถุประสงค์ของการทดสอบส่วนใหญ่ จะใช้เพื่อวินิจฉัยหรือสืบค้นให้ทราบถึงโรคของผู้ป่วย แต่บางครั้งจะใช้เพื่อวินิจฉัยแยกโรคนั้นออกไป ประโยชน์ของวิธีทดสอบจะเกิดแก่ผู้รับการทดสอบหรือแก่ชุมชนนั้น เมื่อโรคดังกล่าวมีลักษณะดังนี้

- 4.1 พบบ่อยหรือมีความชุกสูง
 - 4.2 มีอัตราตายหรือความพิการสูง ถ้าไม่ได้รับการรักษา
 - 4.3 มีวิธีการรักษาให้ผู้ป่วยรอดพ้นจากความทุกข์ทรมานได้ดีเมื่อวินิจฉัยได้แล้ว
 - 4.4 ช่วยลดความพิการหรืออัตราตาย ถ้าวินิจฉัยได้เร็วขึ้น
- ความชุก (prevalence) ของโรคในสังคมแห่งหนึ่ง จะกระทบกระเทือนถึงการเลือกใช้และการแปลผลการทดสอบเพื่อวินิจฉัยโรคดังจะได้แสดงต่อไป ในที่นี้จะเน้นให้เห็นก่อนว่าการนำการทดสอบวิธีหนึ่งไปใช้สถานการณ์แห่งหนึ่ง ควรคำนึงถึงความชุกของโรคนั้นไว้ด้วย ความชุกของโรคในสถานการณ์ที่อ้างถึงดังต่อไปนี้ จะเรียงตามลำดับจากต่ำไปสูงดังนี้

ก. เมื่อใช้ในชุมชนเพื่อสืบหาผู้ป่วยในชุมชนที่ไม่ได้ใช้ประวัติและการตรวจร่างกายเพื่อแยกคนปกติออกไปก่อน วิธีใช้เป็น screening test เช่น การออกหน่วยแพทย์ไปบริการชุมชน

วิธีทดสอบจะต้องไวพอ ใช้ง่าย ราคาถูก แต่ในกรณีนี้ความซุกของโรคที่ต้องการค้นหาจะมีน้อยที่สุด

ข. เมื่อใช้กับผู้ป่วยที่มารับการรักษาด้วยโรคใดๆ ที่โรงพยาบาลหรือคลินิก วิธีนี้ใช้เพื่อค้นหาผู้ป่วยแบบ case finding คนปกติมักไม่มาโรงพยาบาล จึงทำให้ความซุกของโรคต่างๆ สูงกว่าในข้อ 1 เพราะทำการตรวจสอบเฉพาะผู้ป่วยที่มาโรงพยาบาลเท่านั้น

ค. เมื่อใช้กับผู้ป่วยที่แพทย์สงสัยว่าจะป่วยเป็นโรคนั้น วิธีทดสอบที่เลือกใช้ก็เป็นวิธีที่ใช้กับผู้ป่วยในโรงพยาบาล ซึ่งแบ่งแยกผู้ป่วยตามระบบอวัยวะหรือตามหน่วยงาน การวินิจฉัยหรือแยกโรคแบบนี้เป็น การวินิจฉัยโรคในผู้ป่วยที่มีลักษณะคลินิกเฉพาะอย่าง แต่ยังไม่สามารถแยกโรคออกจากกันได้ชัดเจน การประเมินคุณค่าของวิธีการวินิจฉัย (การทดสอบที่กล่าวถึงในบทความนี้จะเกี่ยวข้องกับวิธีการวินิจฉัยที่ใช้ในผู้ป่วยกลุ่มนี้)

ง. เมื่อใช้วินิจฉัยผู้ป่วยที่เป็นโรคนั้นอยู่แล้ว และใช้เพื่อติดตามผลการรักษามากกว่าจะนำมาใช้เพื่อวินิจฉัยโรค เช่น การวัดระดับ thyroxin ในผู้ป่วยที่กำลังรับประทานยา thyroxin เป็นต้น ความซุกของโรคในกรณีนี้มีร้อยละ 100

5. ขั้นตอนการคำนวณค่าต่างๆ ซึ่งเป็นคุณสมบัติของวิธีทดสอบเพื่อวินิจฉัยโรค

คุณสมบัติของวิธีทดสอบเพื่อวินิจฉัยโรค จะประกอบด้วย

5.1 เกณฑ์มาตรฐานที่ใช้แบ่งผู้ที่เป็นโรคที่นำสนใจออกจากผู้อื่น

5.2 จุดตัด (cut-off point) ของผลการทดสอบ เพื่อแบ่งผลลัพธ์ออกเป็นผลบวกและผลลบ หรือช่วงค่าปกติที่แยกค่าผิดปกติออกจากค่าปกติ

5.3 ค่าต่างๆ ที่คำนวณได้จากตาราง 2×2 ซึ่งได้แก่ sensitivity, specificity, positive predictive value, negative predictive value, accuracy, prevalence of the disease

5.1 เกณฑ์มาตรฐาน (gold standard) เกณฑ์มาตรฐานหรือ gold standard มีความสำคัญมาก เพราะเป็นเกณฑ์อ้างอิงที่ทุกคนใช้ในการวินิจฉัยโรค ทำให้เราสามารถแบ่งผู้เข้ารับการทดสอบออกเป็น 2 กลุ่ม คือ กลุ่มมีโรคและไม่มีโรค เกณฑ์มาตรฐานจึงต้องมีเกณฑ์ถูกต้องที่สุด ซึ่งมักจะเป็นการตรวจชิ้นเนื้อ, ผลผ่าตัดการตรวจศพ, การติดตามผู้ป่วยนานๆ จนทราบแน่ชัดว่าผู้ป่วยหรือไม่ได้ป่วยเป็นโรคนั้นๆ โรคที่หาเกณฑ์มาตรฐานได้ยากหรือทำตามเกณฑ์มาตรฐานได้ยาก เช่น เยื่อหุ้มสมองอักเสบจากเชื้อวัณโรคจะทำให้การประเมินคุณค่าของการทดสอบวิธีใหม่ๆ เป็นไปได้ยาก อย่างไรก็ตามโรคที่ยังหาเกณฑ์มาตรฐานได้ยากมักจะเป็นโรคที่มีผู้คิดค้นวิธีการทดสอบใหม่ๆ เพื่อวินิจฉัยโรคให้ถูกต้องมากขึ้น ในกรณีนี้เราต้องพยายามจัดหาเกณฑ์มาตรฐานที่ดีที่สุดและเหมาะสมที่สุดในขณะนั้น เพื่อใช้เป็นหลักในการวินิจฉัยผู้ป่วย เกณฑ์มาตรฐานที่ใช้จึงต้องบอกรายละเอียดเสมอ โดยกำหนดกฎเกณฑ์ไว้ก่อนที่

จะทำการศึกษา ข้อที่พึงระวังสำหรับโรคที่ยังไม่มีเกณฑ์มาตรฐานที่เหมาะสมก็คือ การทดสอบวิธีใหม่ที่ถูกนำมาประเมิน อาจจะทำให้ผลถูกต้องใกล้เคียงกับความเป็นจริงมากกว่าเกณฑ์มาตรฐานที่ใช้อยู่ก็ได้ ซึ่งในกรณีนี้ผลการทดสอบวิธีใหม่นี้จะกลายเป็นเกณฑ์มาตรฐานต่อไปในอนาคต

5.2 การกำหนดจุดตัด (cut-off point) ระหว่างค่าปกติและค่าผิดปกติ

วิธีที่ 1 เป็นการหาค่าปกติจากประชากรกลุ่มใหญ่กลุ่มเดียว โดยยึดหลักว่าประชากรกลุ่มนี้เป็นกลุ่มปกติไว้ก่อน การหาค่าปกติในกรณีนี้กระทำโดยวัดค่าตัวแปรและนำค่าดังกล่าวมาแจกแจงเป็นฮิสโตแกรม หรือลากเป็นเส้นโค้งเพื่อดูลักษณะการกระจายของค่าตัวแปรที่วัดได้ จากนั้นสามารถกำหนดจุดตัดตามที่ต้องการดังนี้

5.2.1 ถ้าการกระจายของค่าตัวแปรเป็นการแจกแจงแบบปกติ (normal distribution) ก็สามารถหาค่าเฉลี่ยและพิสัยของค่าที่ได้จากประชากรส่วนใหญ่ได้ (ส่วนใหญ่คือร้อยละ 95.4) โดยกำหนดว่าค่าที่อยู่ใกล้กับค่าเฉลี่ยและพบในประชากรร้อยละ 95.4 เป็นค่าปกติ ค่าที่ได้จากประชากรที่เหลือประมาณร้อยละ 5 ซึ่งเป็นค่าที่อยู่ห่างกว่าค่าเฉลี่ยจะเป็นค่าผิดปกติ ซึ่งอาจจะแบ่งเป็นค่าที่ต่ำและสูงเกินไปอย่างละร้อยละ 2.5 ก็ได้ การศึกษาหาค่าปกติโดยวิธีนี้สะดวกรวดเร็ว แต่จุดอ่อนอยู่ที่นำประชากรกลุ่มเดียวมาศึกษาทำให้ไม่ทราบว่ามีค่าสูงหรือต่ำเกินไปจะเป็นคนผิดปกติหรือป่วยเป็นโรคจริงหรือไม่ ยกตัวอย่าง เช่น ผู้ที่มีไอคิวสูงจะจัดเป็นผู้ป่วยได้ไหม? นอกจากนี้การหาค่าปกติจากวิธีนี้จะทำให้โรคต่างๆ มีความซุกเหมือนกันหมด คือ เท่ากับร้อยละ 5

5.2.2 ถ้าการกระจายของตัวแปรเบี่ยงไปข้างใดข้างหนึ่ง ก็อาจจะดัดแปลงให้เป็นการแจกแจงแบบปกติได้โดยใช้หลักคณิตศาสตร์เข้าช่วยแล้วใช้วิธีการในข้อ 5.2.1 ถ้าไม่ดัดแปลงก็สามารถใช้วิธีหา percentile แทน คือ ใช้ percentile 50 เป็นค่าเฉลี่ย และพิสัยอยู่ระหว่าง percentile 2.5-97.5 ซึ่งใช้เป็นจุดตัดด้วย ถ้าโรคนี้มีแต่ค่าที่สูงเกินไปหรือต่ำเกินไป ก็สามารถใช้ percentile ที่ 5.0 หรือ 95.0 มาเป็นจุดตัดก็ได้ ข้อดีและจุดอ่อนจะเหมือนกับในหัวข้อ 5.2.1 ดังกล่าวแล้ว วิธีนี้ดีกว่า 5.2.1 ที่การแจกแจงของตัวแปรไม่ต้องเป็นแบบปกติ

วิธีการหาค่าปกติในข้อ 5.2.1 และ 5.2.2 อาศัยหลักที่ว่า การกระจายของตัวแปรจะอยู่ใกล้เคียงกัน ถ้าได้มาจากประชากรกลุ่มเดียวกัน (ประชากรที่มีลักษณะเฉพาะเดียวกัน) ที่มีจำนวนมากพอทำให้เกิดเป็นการแจกแจงแบบ unimodel ค่าปกติที่ได้จากวิธีทั้ง 2 นี้เป็นการศึกษาแบบพรรณนา ซึ่งบอกถึงลักษณะการกระจายของตัวแปรและค่าเฉลี่ยในกลุ่มประชากรปกติ โดยไม่ต้องเปรียบเทียบกับประชากรกลุ่มอื่น จึงทำให้ผู้วิจัยไม่เห็นภาพพจน์ของค่าตัวแปรที่เหลื่อมล้ำกันระหว่างคนปกติและผู้ป่วยซึ่งเกิดขึ้นเสมอ ถ้าค่าของตัวแปรที่ได้จากคนปกติไม่เหลื่อมล้ำกับค่าที่ได้จากผู้ป่วย

วิธีทดสอบที่ใช้วัดค่าตัวแปรนี้จะกลายเป็นวิธีอุดมคติซึ่งดีที่สุด และวิธีการหาจุดตัดทั้ง 2 แบบจะสะดวกมาก

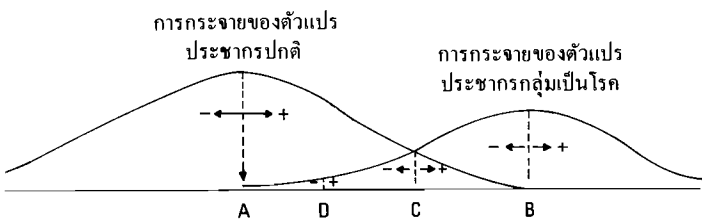
วิธีที่ 2 เป็นการหาจุดตัดที่เหมาะสมระหว่างกลุ่มประชากรที่เป็นโรคเปรียบเทียบกับกลุ่มประชากรปกติ หรือกลุ่มประชากรที่ไม่ได้เป็นโรคนั้นแต่มีลักษณะคลินิกเหมือนกัน เมื่อวัดค่าตัวแปรที่ต้องการได้แล้วจะนำค่าที่ได้จากประชากรกลุ่มต่างๆ มาเปรียบเทียบการกระจายและการแจกแจงความถี่บนฐานเดียวกัน **วิธีนี้ยึดประชากรกลุ่มที่เป็นโรคเป็นหลัก** แล้วนำประชากรกลุ่มอื่นมาเปรียบเทียบ

5.2.3 ถ้าใช้วิธีทดสอบนี้ในสถานพยาบาล ซึ่งมีแต่ผู้ป่วยมารับการทดสอบลักษณะคลินิกของผู้ป่วยคล้ายคลึงกัน แต่เกิดจากสาเหตุต่างกันและเป็นคนละโรค วิธีทดสอบที่นำมาใช้ก็เพื่อที่จะแยกผู้ป่วยโรคหนึ่งออกจากอีกโรคหนึ่ง การศึกษาหาจุดตัดในสถานการณ์เช่นนี้ให้ใช้กลุ่มเปรียบเทียบเป็นประชากรกลุ่มที่มีลักษณะคล้ายคลึงกับประชากรกลุ่มที่เป็นโรค

5.2.4 ถ้าใช้วิธีการทดสอบนี้ในประชากรทั่วไปเพื่อค้นหาโรคใดโรคหนึ่ง ให้ใช้กลุ่มเปรียบเทียบเป็นกลุ่มประชากรปกติ หรือจะนำกลุ่มประชากรที่มีลักษณะคล้ายคลึงกัน แต่มีสาเหตุแตกต่างจากประชากรกลุ่มเป็นโรคมาร่วมพิจารณาด้วย โดยรวมกลุ่มประชากรปกติและกลุ่มที่มีลักษณะคล้ายคลึงกันมาเป็นกลุ่ม (เปรียบเทียบ) เดียวกันก็ได้

หลังจากที่แสดงการกระจายและการแจกแจงของค่าตัวแปรจากประชากรกลุ่มต่างๆ บนฐานเดียวกันแล้ว จะพบว่ามีค่าเฉลี่ยต่างกันเสมอระหว่างประชากรกลุ่มต่างๆ การพิจารณาจุดตัดที่เหมาะสมจึงต้องคำนึงถึงผลดีและผลเสียของกลุ่มที่ให้ผลบวกจริงและเท็จ และกลุ่มที่ให้ผลลบจริงและเท็จอีกด้วย ดังที่จะแสดงต่อไป

การกำหนดจุดตัดของค่าตัวแปรเพื่อแบ่งผลการทดสอบออกเป็นบวกและลบมีความสำคัญ อย่างไรก็ตามการเลือกจุดตัดจะต้องพิจารณาหลายๆ ด้านและจะมีหลายจุดให้เลือกเป็นจุดตัดดังที่จะแสดงต่อไป



ภาพนี้แสดงการกระจายของค่าตัวแปรที่เหลื่อมล้ำกันระหว่างประชากร 2 กลุ่ม และการเลือกจุดตัด A, B, C และ D เพื่อแบ่งผลการทดสอบเป็นบวกและลบตามจุดต่างๆ (ค่าตัวแปรที่วัดได้เป็น continuous variable)

จุดตัด A เป็นจุดตัดที่จะทำให้ประชากรกลุ่มเป็นโรคทุกรายให้ผลบวกกับการทดสอบ การกำหนดแบบนี้จะให้ค่า sensitivity สูงและเกิดผลบวกเท็จได้มาก การกำหนดแบบนี้จึงเหมาะสำหรับ

กรณีที่มีโรคนั้นมีผลร้ายรุนแรง แต่รักษาให้หายขาดหรือป้องกันผลร้ายแรงที่จะเกิดขึ้นในอนาคตได้ ขณะเดียวกันการรักษาผู้ที่ให้ผลบวกเท็จก็ไม่เกิดผลร้าย หรือมีวิธีทดสอบต่ออีก 1 วิธีที่จะมากกว่าการตรวจให้ผลบวกเท็จออกไปจากผู้ที่เป็นโรคจริง เช่น การตรวจปฏิกิริยา VDRL, การถ่ายภาพจากระแหวนเพียง 1 ครั้งในระยะที่มีการระบาดของหิวขาดโรค ข้อเสียของการกำหนดจุดนี้อยู่ที่จำนวนผู้ที่ให้ผลบวกเท็จซึ่งถูกวินิจฉัยว่าเป็นโรค เพราะจะเกิดความหวาดกลัวจากการที่ถูกหลอกว่าเป็นโรคและอาจจะได้รับพิษจากยาที่รักษาโรค จึงต้องพิจารณาให้รอบคอบก่อนจะเลือกจุดตัดนี้

จุดตัด B เป็นจุดตัดที่ประชากรผู้ไม่เป็นโรคทุกรายให้ผลลบ ส่วนประชากรที่เป็นโรคบางรายจะวินิจฉัยถูกว่าเป็นโรค แต่อีกส่วนหนึ่งจะให้ผลลบเท็จและไม่ถูกวินิจฉัยว่าเป็นโรค จุดตัดนี้เหมาะสำหรับกรณีที่มีโรคนั้นเป็นโรคเรื้อรังที่มีลักษณะดังนี้ ไม่มีวิธีรักษาให้หายขาด ต้องรักษานาน การรักษาต้องเสี่ยงต่อผลร้ายจากยา การรักษามีผลเล็กน้อยในการเปลี่ยนแปลงการดำเนินโรค. ในกรณีนี้ถึงแม้ผู้ที่ให้ผลลบเท็จไม่ได้รับการรักษาก็ไม่เกิดผลเสียหายร้ายแรงและไม่แพร่โรคให้แก่ผู้อื่น โรคกลุ่มนี้ได้แก่ โรคทาง collagen vascular disease และมะเร็งเม็ดเลือดขาวในผู้ใหญ่ที่ยังไม่มีวิธีรักษาให้หายขาด เป็นต้น

จุดตัด C เป็นจุดตัดแยกกลุ่มประชากรที่เป็นโรคและไม่เป็นโรคออกจากกันโดยมีจำนวนผู้ที่ให้ผลบวกเท็จและผลลบเท็จน้อยที่สุด การกำหนดวิธีนี้เป็นแบบที่นิยมเพราะมีความถูกต้องมาก ผู้ที่เป็นโรค (ผลบวกจริง) จะได้รับการรักษา ส่วนผู้ที่ไม่เป็นโรคแต่ได้รับยาจะเกิดผลร้ายน้อย หรือไม่มีเลยจากผลบวกเท็จ ผลเสียที่เกิดขึ้นก็จะมีน้อยด้วยในกรณีผู้ที่เป็นโรคแต่ให้ผลลบเท็จ โดยสรุปแล้วผลเสียที่เกิดจากผลบวกเท็จและผลลบเท็จมีขนาดใกล้เคียงกัน ในทางปฏิบัติอาจจะคำนวณจุดตัดนี้ได้จากการรวมค่าของ sensitivity และ specificity ของแต่ละจุดตัดเพื่อดูว่าค่าใดจะให้ค่าสูงสุด หรือจะใช้ลักษณะของ receiver operating characteristic (ROC) curve มาเลือกจุดตัดก็ได้ ดังตัวอย่างที่แสดงไว้ในตอนหลังของบทความ การที่ไม่ได้กล่าวถึงตอนนี้เพราะจะต้องทราบคุณสมบัติอื่นๆ ของวิธีทดสอบก่อน ซึ่งต้องนำไปใช้ในการสร้าง ROC curve ด้วย

จุดตัด D เป็นจุดล่อยที่ใช้หลักเกณฑ์อื่นๆ มาตัดสิน เช่น การรักษาที่เริ่มจากจุดนี้จะให้ประโยชน์สูงสุด เป็นต้น

จุดตัดต่างๆ สำหรับโรคเดียวกันอาจจะมีมากกว่า 1 จุดก็ได้ ถ้ามีวัตถุประสงค์ต่างกัน เช่น ในเรื่องโรคเอดส์ ผู้ที่เป็นโรคจริงเท่านั้นที่สมควรได้รับการวินิจฉัยว่าเป็นโรคนี จุดตัดจึงใช้จุด B ขณะเดียวกันในด้านการป้องกันการระบาดของโรคอาจจะแนะนำผู้ที่ได้ค่าการทดสอบระหว่างจุด A และ B ระมัดระวังในการแพร่เชื้อแก่ผู้อื่นโดยทางเพศสัมพันธ์ นอกจากนี้การใช้วิธีทดสอบอื่นๆ มาร่วมตรวจวัดจะช่วยแยกผู้ป่วยหรือผู้ที่ไม่ป่วยอีกจำนวนหนึ่งออกจากกลุ่มที่เคยให้ผลลบเท็จหรือผลบวกเท็จจากการทดสอบที่ผ่านมา

โดยทั่วไปถ้าต้องการตรวจค้นโรคที่รักษาให้หายได้จากกลุ่มประชากร จะเลือกวิธีการทดสอบที่มีความไวสูงมากก่อน แล้วจึงนำผู้ที่ให้ผลบวกจากการทดสอบครั้งแรกมาทดสอบกับวิธีที่ 2 ซึ่งมีค่าตรวจจำเพาะสูง ทั้งนี้เพื่อแยกผู้ที่ให้ผลบวกเท็จออกไป

5.3 การสร้างและแปลผลการทดสอบจากตาราง 2 x 2 ขึ้นตอนที่ง่ายที่สุดในการคำนวณหาค่าต่างๆ ซึ่งเป็นคุณสมบัติของวิธีทดสอบใดๆ คือ การสร้างตารางที่เรียกว่าตาราง 2 x 2 (two by two table) โดยทฤษฎีแล้ววิธีทดสอบที่อุดมคติเพื่อวินิจฉัยโรคใดโรคหนึ่ง คือ วิธีทดสอบที่ผู้ป่วยทุกรายให้ผลบวกเมื่อป่วยเป็นโรคนั้น คือ มี sensitivity = 100% และผู้อื่นทุกรายที่ไม่ได้ป่วยเป็นโรคนั้นให้ผลลบ คือ มี specificity = 100% (อาจจะเป็นคนปกติหรือป่วยเป็นโรคอื่นที่มีลักษณะคลินิกลคล้ายคลึงกับโรคนั้น) การแบ่งกลุ่มประชากรออกเป็น 2 กลุ่ม คือ กลุ่มที่เป็นโรคและไม่ได้เป็นโรคนั้นอาศัยเกณฑ์มาตรฐานที่เรียกว่า gold standard มาตัดสิน เช่น ผู้ป่วยมะเร็งเต้านมก็ต้องอาศัยผลการตรวจชิ้นเนื้อ ผู้ป่วยติดเชื้อทรวงแท้แกรมลบในเลือด (septicemia) ก็ต้องอาศัยผลการเพาะเชื้อจากเลือด เป็นต้น เกณฑ์มาตรฐานนี้ต้องไม่รวมผลของวิธีการทดสอบที่กำลังจะประเมินคุณค่าอยู่ด้วย เกณฑ์มาตรฐานจะแบ่งผู้ป่วย (และคนปกติหรือผู้ป่วยโรคอื่น) ออกเป็น 2 กลุ่ม คือ กลุ่มมีโรค (นั้น) และกลุ่มไม่มีโรค (นั้น) ซึ่งจะย่อว่า มีโรค และ ไม่มีโรค ดังแผนภูมิที่สร้างขึ้น ดังนี้

เกณฑ์มาตรฐาน (Gold Standard)

มีโรค	ไม่มีโรค
-------	----------

ถ้ามีวิธีทดสอบที่เป็นอุดมคติจะสร้างตารางที่มีเพียง 2 ช่อง ดังนี้

เกณฑ์มาตรฐาน (Gold standard)

	มีโรค	ไม่มีโรค
ผลการทดสอบ	ให้แต่ผลบวกซึ่งเป็นผลบวกจริง (true positive)	ให้แต่ผลลบซึ่งเป็นผลลบจริง (true negative)

แต่เนื่องจากไม่มีวิธีทดสอบเพื่อวินิจฉัยโรควิธีใดเป็นไปตามอุดมคติ กล่าวคือ กลุ่มที่มีโรคบางรายก็ให้ผลการทดสอบเป็นลบ ทั้งๆ ที่ป่วยเป็นโรคนั้น (false negative) และผู้ป่วยโรคอื่นบางรายก็ให้ผลบวกทั้งๆ ที่ไม่ได้เป็นโรคนั้น (false positive) จึงต้องใช้วิธีการตรวจการทดสอบที่ง่ายและเข้าใจได้เร็ว คือ การสร้างตาราง 2 x 2 หรือเรียกว่า two by two table ดังนี้

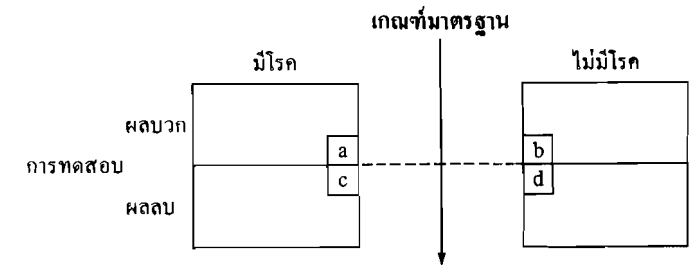
เกณฑ์มาตรฐาน (Gold standard)

	มีโรค	ไม่มีโรค
ให้ผลบวก	ผลบวกจริง (true positive)	ผลบวกเท็จ (false positive)
ผลการทดสอบ	a b	
ให้ผลลบ	ผลลบเท็จ (false negative)	ผลลบจริง (true negative)
	c d	

ตัวอักษรประจำช่องคือ a, b, c, และ d หมายถึง จำนวนผู้ป่วยในแต่ละช่อง ของตาราง 2 x 2 ที่ให้ผลการทดสอบตามช่องตาราง

จากตารางดังกล่าว เราจะเข้าใจความหมายของค่าต่างๆ ที่เป็นคุณสมบัติของการทดสอบได้ดังต่อไปนี้

ถ้ามองตามแนวดิ่ง ซึ่งแบ่งผู้ป่วยออกเป็นกลุ่มมีโรคและไม่มีโรค ดังแผนภูมิข้างล่างนี้



ผลการมองตามแนวดิ่งจะเกิดคำว่า sensitivity และ specificity

sensitivity หมายถึงจำนวนผู้ป่วยคิดเป็นร้อยละที่ให้ผลบวก โดยการทดสอบวิธีนั้น เมื่อป่วยเป็นโรคนั้นจริงตามเกณฑ์มาตรฐานที่จุดตัดนั้น

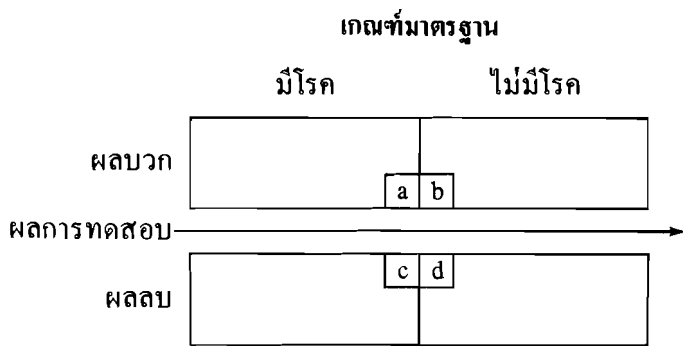
วิธีคำนวณค่า sensitivity ของการทดสอบคิดเป็นร้อยละ = $\frac{a}{a+c} \times 100\%$

Specificity หมายถึงจำนวนผู้ป่วยคิดเป็นร้อยละที่ให้ผลลบ โดยการทดสอบวิธีนั้น เมื่อไม่ได้ป่วยเป็นโรคนั้นตามเกณฑ์มาตรฐานที่จุดตัดนั้น

วิธีคำนวณค่า specificity ของการทดสอบคิดเป็นร้อยละ = $\frac{d}{b+d} \times 100\%$

จะเห็นว่า sensitivity (ความไว) เกี่ยวข้องกับกลุ่มที่มีโรค ในขณะที่ specificity (ความจำเพาะ) เกี่ยวข้องกับกลุ่มที่ไม่มีโรค ค่าของ sensitivity และ specificity ของวิธีทดสอบใดๆ ที่จุดตัดจุดหนึ่ง โดยทั่วไปถือว่า มีค่าคงที่ คือเป็นคุณสมบัติประจำตัวของวิธีทดสอบนั้น นอกจากนี้มีการเปลี่ยนแปลงเกณฑ์มาตรฐานที่ใช้ตัดสินว่ามีหรือไม่มีโรค หรือมีการเปลี่ยนจุดตัดแยกค่าบวกและลบ

ถ้ามองตามแนวนอน ซึ่งแบ่งผู้ป่วยออกเป็น 2 ข้าง ตามผลบวกหรือผลลบจากการทดสอบ



ผลการมองในแนวนอนนี้จะได้ค่า positive และ negative predictive values ซึ่งสามารถคำนวณได้ตามวิธีที่แสดงดังนี้

Positive predictive value หมายถึงจำนวนประชากรที่เป็นโรค คิดเป็นร้อยละ *ป่วยเป็นโรคนั้น* จริง ตามเกณฑ์มาตรฐาน ในบรรดาผู้ที่ให้ผลบวกจากการทดสอบ

เราสามารถคำนวณค่า positive predictive value คิดเป็นร้อยละได้โดย
$$= \frac{a}{a + b} \times 100\%$$

Negative predictive value หมายถึงจำนวนประชากรคิดเป็นร้อยละ *ที่ไม่ได้ป่วยเป็นโรคนั้น* ตามเกณฑ์มาตรฐานในบรรดาผู้ที่ให้ *ผลลบ* จากการทดสอบ

เราสามารถคำนวณค่า negative predictive value ได้โดย
$$= \frac{d}{c + d} \times 100\%$$

ค่าของ positive และ negative predictive values *เปลี่ยนแปลงได้มาก* ตามความชุกของโรคในกลุ่มคนที่ถูกทดสอบด้วยวิธีนี้ ดังที่ได้แสดงในตารางที่ 1 ในทางปฏิบัติแพทย์สั่งให้มีการทดสอบเมื่อมีปัญหาในการวินิจฉัยโรค ซึ่งแสดงว่าในขณะนั้นแพทย์ยังวินิจฉัยโรคที่แน่นอนไม่ได้ และต้องการข้อมูลเพิ่มเติมจากผู้ป่วย

ปัญหาที่ตามมาคือการแปลผลการทดสอบ นั่นคือเมื่อการทดสอบชนิดหนึ่งให้ผลบวก ผลบวกที่ได้น่าจะเป็นผลบวกจริงหรือเท็จ เช่นเดียวกับถ้าให้ผลลบน่าจะเป็นผลลบจริงหรือเท็จ การแปลผลให้ถูกต้องจะต้องทราบความชุกของโรคนั้นในประชากรที่เรา กำลังศึกษา ดังจะได้แสดงต่อไป

ถ้าเป็นโรคที่หายาก คือ มีความชุกของโรคต่ำมากและมี specificity ของวิธีการทดสอบต่ำเมื่อการทดสอบให้ผลบวก เรามีแนวโน้มที่จะแปลว่าเป็นผลบวกเท็จ นั่นคือ positive predictive value ของการทดสอบเพื่อวินิจฉัยโรคในชุมชนนั้นมีค่าต่ำมาก แต่ถ้าได้ผลลบน่าจะเป็นผลลบจริง โดยเฉพาะอย่างยิ่งถ้า sensitivity ของการทดสอบวิธีนี้สูงมาก นั่นคือ negative predictive value ของการทดสอบเพื่อวินิจฉัยโรคในชุมชนนั้นมีค่าสูง ให้ทดลองสมมติค่าต่างๆ ในตาราง 2 x 2 แล้วทำการคำนวณ positive และ negative predictive values ก็ จะเห็นความจริงข้อนี้ชัดเจนขึ้น

การมองในแนวนอนนี้จึงเหมือนกับสภาพทางคลินิกที่แพทย์ ได้รับรายงานผลการทดสอบอย่างหนึ่งแล้ว เช่น ทราบผลการทดสอบ

cold agglutinin แล้ว และอยากทราบต่อไปอีกว่าถ้าผู้ป่วยรายนี้ให้ผลบวกหรือลบกับการทดสอบ cold agglutinin ผู้ป่วยมีความน่าจะเป็นโรคติดเชื้อมัยโคพลาสมาเพียงใด การแปลผลการทดสอบทางคลินิกก็คือ การใช้ positive และ negative predictive values นั้นเอง และในขณะนั้นผู้รักษายังไม่ทราบผลการทดสอบที่เป็น gold standard

ถ้ามองดูความถูกต้องของผลการทดสอบวิธีนี้ (accuracy of the test) กล่าวคือ เมื่อผู้ป่วยมีโรคนั้นจริงต้องให้ผลบวก (true positive) และเมื่อผู้ป่วยไม่มีโรคนั้นต้องให้ผลลบ (true negative) เราสามารถคำนวณหาค่าความแม่นยำของวิธีทดสอบคิดเป็นร้อยละได้โดย
$$= \frac{a + d}{a + b + c + d} \times 100\%$$
 จากสูตรที่ใช้จะทำให้ทราบว่า accuracy หมายถึงจำนวนผู้ป่วยส่วนหนึ่งคิดเป็นร้อยละของบรรดาผู้ป่วยทั้งหมดที่ทำการทดสอบและให้ผลบวกเมื่อมีโรค หรือให้ผลลบเมื่อไม่มีโรคตามเกณฑ์มาตรฐาน จะสังเกตเห็นว่าวิธีทดสอบที่อุดมคติจะได้ค่าของ accuracy = 100%

ส่วนความชุกของโรคในการศึกษาครั้งนี้ (prevalence or pretest probability of a disease) จะคำนวณได้จาก
$$= \frac{a + c}{a + b + c + d}$$
 (ให้คุณด้วย 100 ถ้าคิดเป็นร้อยละ)

แผนภูมิและตารางที่จะแสดงต่อไปนี้ เป็นการสรุปและทบทวนคุณสมบัติ รวมทั้งศัพท์ต่างๆ ที่ใช้ในตาราง 2 x 2 และแสดงให้เห็นผลของความชุกของโรคต่อการแปลผลการทดสอบ โดยกำหนดค่า sensitivity เท่ากับร้อยละ 85 และ specificity เท่ากับร้อยละ 90 ส่วนคำว่า positive predictive value มีความหมายเหมือน predictive value of a positive test ทำนองเดียวกันคำว่า negative predictive value มีความหมายเหมือน predictive value of a negative test เช่นกัน

Fourfold table demonstrating "blind" comparison with "gold standard"

		Gold standard		
		Patient has the disease	Patient does not have the disease	
Test result (conclusion drawn from the results of the test)	Positive Patient appears to have the disease	True positive	False positive	a + b
	Negative Patient appears not to have the disease	False negative	True negative	c + d
		a + c	b + d	a + b + c + d

Stable properties:
a/(a+c) = sensitivity
d/(b+d) = specificity

Frequency-dependent properties:
a/(a+b) = positive predictive value
d/(c+d) = negative predictive value
(a+d)/(a+b+c+d) = accuracy
(a+c)/(a+b+c+d) = prevalence

ตารางที่ 1 แสดงผลของความชุกของโรคต่อการแปลผลการทดสอบเมื่อให้ sensitivity = ร้อยละ 85 และ specificity = ร้อยละ 90 ทุกครั้ง

Prevalence (Pre-test likelihood or Prior probability of disease)	99%	95%	90%	80%	70%	60%	50%	40%	30%	20%	10%	6%	1%	0.5%	0.1%
Predictive Value of a Positive test (Post-test likelihood or posterior probability of disease following a Positive test result)	99.9%	99.4%	99%	97%	95%	93%	89%	85%	78%	68%	48%	11%	7%	4%	1%
Predictive value of a Negative test (Post-test likelihood or posterior Probability of NO Disease following a Negative test result)	6%	24%	40%	60%	72%	80%	86%	90%	93%	96%	98%	99%	99.8%	99.9%	99.98%
Post-test likelihood or Posterior probability OF disease following a Negative test result	94%	76%	60%	40%	28%	20%	14%	10%	7%	4%	2%	1%	0.2%	0.1%	0.02%

6. ตัวอย่างแสดงการคำนวณค่าต่างๆ ในการทดสอบเพื่อวินิจฉัยโรค

การตรวจหา cold agglutinin ข้างเดียว เป็นวิธีที่ให้ผลในเวลา 5 นาทีทำได้รวดเร็ว และมีรายงานบางฉบับเสนอให้นำมาใช้ทางคลินิก เพื่อวินิจฉัยผู้ป่วยที่ติดเชื้อมัยโคพลาสมาในระบบทางเดินหายใจส่วนล่าง เพราะการเพาะเชื้อมัยโคพลาสมาเป็นวิธีที่ยุ่งยากและยังสามารถพบเชื้อในผู้ที่ไม่มีอาการใดๆ ส่วนการตรวจทางน้ำเหลืองโดยใช้ CF test ก็ต้องเจาะเลือด 2 ครั้ง ห่างกัน 2 สัปดาห์กว่าจะทราบผล การศึกษาเรื่องนี้ในโรงพยาบาลแห่งหนึ่งพบว่าการตรวจหา cold agglutinin ข้างเดียวในผู้ป่วยที่รับไว้ ให้ผลบวก 11 รายในผู้ป่วย 12 รายที่ติดเชื้อมัยโคพลาสมา และให้ผลบวกในผู้ป่วยโรคอื่นๆ อีก 13 รายจาก 40 ราย การตรวจหา cold agglutinin ข้างเดียวให้ผลบวกเมื่อผู้ป่วยมีไตเตอร์ของ cold agglutinin ซึ่งตรวจโดยวิธีมาตรฐานเท่ากับหรือสูงกว่า 1:64 ผลการศึกษาดังกล่าวสามารถแสดงในตารางดังนี้

ตารางแสดงคุณค่าของการทดสอบหา cold agglutinin ข้างเดียวในการวินิจฉัยโรคติดเชื้อมัยโคพลาสมาโดยใช้ตาราง 2 x 2

		โรคติดเชื้อมัยโคพลาสมา	
		มีโรค	ไม่มีโรค
การทดสอบหา cold agglutinin ข้างเดียว	ให้ผลบวก	11	13
	ให้ผลลบ	1	27

จากค่าต่างๆ ที่แสดงในตาราง 2 x 2 เราสามารถคำนวณค่าต่างๆ ได้ดังนี้

$$\begin{aligned} \text{Sensitivity} &= \frac{11}{11+1} \times 100 = 91.7\% \\ \text{Specificity} &= \frac{27}{13+27} \times 100 = 67.5\% \\ \text{Positive predictive value} &= \frac{11}{11+13} \times 100 = 45.8\% \\ \text{Negative predictive value} &= \frac{27}{27+1} \times 100 = 95.4\% \\ \text{Accuracy of the test} &= \frac{11+27}{11+13+1+27} \times 100 = 73.1\% \\ \text{Prevalence of the disease} &= \frac{11+1}{11+13+1+27} \times 100 = 23.1\% \end{aligned}$$

สรุปได้ว่า ความไวและความจำเพาะของการทดสอบหา cold agglutinin ข้างเดียวในการวินิจฉัยโรคติดเชื้อมัยโคพลาสมาในระบบทางเดินหายใจส่วนล่าง มีค่าเท่ากับร้อยละ 91.7 และ 67.5 ตามลำดับ และถ้าผู้ป่วยรายใดให้ผลบวกเราจะแปลผลว่าโอกาสที่จะเป็นโรคนี้นี้เพียงร้อยละ 45.8 แต่ถ้าให้ผลลบเราจะแปลผลว่าโอกาสที่จะไม่เป็นโรคนี้นี้มีค่าสูงถึงร้อยละ 95.4 ความชุกของโรคนี้ในการศึกษามีค่าเท่ากับร้อยละ 23.1

ในกรณีที่ความชุกของโรคนี้เปลี่ยนแปลงไป จะทำให้เราต้องเปลี่ยนการแปลผลบวกและลบด้วย ทั้งนี้เราต้องทราบค่าความชุกที่เปลี่ยนแปลงพร้อมกับความไวและความจำเพาะ ซึ่งเป็นคุณสมบัติคงที่ของวิธีทดสอบด้วย ก็จะทำให้เราสามารถแปลผลบวกและลบได้อย่างถูกต้อง ยกตัวอย่าง เช่น

ถ้านำวิธีทดสอบนี้ไปค้นหาผู้ป่วยติดเชื้อมัยโคพลาสมาในชุมชนแห่งหนึ่ง ซึ่งมีอุบัติการณ์เท่ากับ 3 รายใน 100 ราย ที่มีลักษณะคลินิกของการติดเชื้อในระบบทางเดินหายใจส่วนล่าง และสมมติว่า sensitivity และ specificity ของวิธีนี้ยังเท่ากับร้อยละ 91.7 และ 67.5 ตามลำดับ

สมมติว่ามีผู้ติดเชื้อในระบบทางเดินหายใจส่วนล่างมารับการตรวจ 3,000 รายในชุมชนแห่งนั้น เราสามารถคำนวณค่าต่างๆ ได้ดังนี้

ผู้ป่วยโรคติดเชื้ออหิวาต์มีจำนวน
 $= 3,000 \times \frac{3}{100} = 90$ ราย
 ผู้ป่วยที่ไม่ป่วยเป็นโรคติดเชื้ออหิวาต์มีจำนวน
 $= 3,000 - 90 = 2,910$ ราย
 หรือ $= 3,000 \times \frac{97}{100} = 2,910$ ราย
 เนื่องจาก sensitivity ของวิธีทดสอบ = 91.7%
 จำนวนผู้ป่วยโรคติดเชื้ออหิวาต์ที่ให้ผลบวก
 $= \frac{91.7}{100} \times 90 = 83$ ราย (ปัด 0.53 เป็น 1)
 เพราะฉะนั้นผู้ป่วยโรคติดเชื้ออหิวาต์ที่ให้ผลลบ (เท็จ) มีจำนวน
 $= 90 - 83 = 7$ ราย
 ในทำนองเดียวกันเนื่องจาก specificity = 67.5%
 ผู้ที่ไม่มีโรคนี้จะให้ผลลบ $= \frac{67.5}{100} \times 2,910$
 $= 1,965$ ราย (ปัด 0.25 เป็น 1)
 เพราะฉะนั้นผู้ที่ไม่มีโรคนี้แต่ให้ผลบวก (เท็จ)
 $= 2,910 - 1,965 = 945$ ราย
 ให้เติมค่าต่างๆ ที่คำนวณได้ลงในตาราง 2x2 ดังนี้

		โรคติดเชื้ออหิวาต์ในระบบทางเดินหายใจส่วนล่าง		
		มีโรค	ไม่มีโรค	รวม
การทดสอบ	ให้ผลบวก	83	945	1,028
	ให้ผลลบ	7	1,965	1,972
รวม		90	2,910	3,000

เมื่อได้จำนวนผู้ที่ให้ผลบวกและผลลบแล้ว เราสามารถคำนวณหา positive และ negative predictive values ได้ดังนี้

การคำนวณหา positive predictive value
 $= \frac{83}{1,028} \times 100 = 8.07\%$

การคำนวณหา negative predictive value
 $= \frac{1,965}{1,972} \times 100 = 99.6\%$

ซึ่งแสดงว่า การนำการทดสอบเพื่อวินิจฉัยโรควิธีนี้ไปใช้ในชุมชนซึ่งความชุกของโรคต่ำ ถ้าการทดสอบ cold agglutinin ให้ผลลบ จะมีความถูกต้องในการหาว่าผู้ป่วยไม่เป็นโรคนี้ถึงร้อยละ 99.6

แต่ถ้าการทดสอบ cold agglutinin ให้ผลบวก จะมีความถูกต้องในการหาว่าเป็นผู้ป่วยโรคติดเชื้ออหิวาต์เพียงร้อยละ 8.07 ส่วนอีกร้อยละ 91.93 เป็นผลบวกเท็จ ตัวอย่างนี้จะเห็นได้

ชัดเจนว่า ความชุกของโรคมีผลอย่างมากในการแปลผลบวกและผลลบจากการทดสอบ โดยสรุปแล้วถ้าความชุกของโรคต่ำมาก ผลบวกที่ได้จากการทดสอบมักจะเป็น ผลบวกเท็จ

7. การสร้าง ROC curve โดยใช้จุดตัดหลายจุด

ROC curve เป็นอีกวิธีหนึ่งในการพิจารณาหาจุดตัดที่เหมาะสมของค่าที่วัดได้ เพื่อให้ได้ค่าความไวและความจำเพาะที่เหมาะสมที่สุด นอกจากนี้วิธีนี้ยังแสดงคุณค่าของวิธีการทดสอบว่ามีประโยชน์มากน้อยเพียงใดได้ด้วย การสร้าง ROC curve ต้องอาศัยจุดตัดหลายจุด เพื่อที่จะสร้างกราฟของค่าความไวและความจำเพาะที่เกิดจากจุดตัดแต่ละจุด ดังตัวอย่างที่จะแสดงต่อไปนี้

การวัดค่า CPK ในผู้ป่วยที่มีอาการเจ็บ แน่นหน้าอก ซึ่งประกอบด้วยผู้ป่วยกลุ่มที่เป็นกล้ามเนื้อหัวใจวายและกลุ่มที่ไม่มีกล้ามเนื้อหัวใจวาย แต่เป็นโรคอื่นๆ ได้ผลดังในตารางข้างล่างนี้

ตารางแสดงจำนวนผู้ป่วยแต่ละกลุ่มที่ให้ค่า CPK ในช่วงต่างๆ

		กล้ามเนื้อหัวใจวาย	
		มี	ไม่มี
ค่า CPK	> 280	97	1
	80 - 279	118	15
	40 - 79	13	26
	1 - 39	2	88
		230	130

7.1 การใช้จุดตัดหลายจุดเพื่อกำหนด sensitivity และ specificity ของจุดตัดแต่ละจุด การสร้าง ROC curve จะต้องแบ่งค่า CPK ออกเป็นหลายช่วง โดยใช้จุดตัดหลายจุดแทนที่จะใช้จุดเดียว และสร้างตาราง 2x2 ตามจำนวนจุดตัดที่เลือกใช้ ประโยชน์ของการใช้จุดตัดหลายจุดยังมีอีกซึ่งจะได้กล่าวถึงในเรื่อง likelihood ratio สำหรับข้อมูล CPK จากผู้ป่วยกล้ามเนื้อหัวใจวาย เราจะเลือกค่า CPK มากกว่าหรือเท่ากับ 280, 80, 40 และ 0 เป็นจุดตัดและสร้างตาราง 2x2 ได้ดังนี้

เมื่อใช้จุดตัดของ CPK มากกว่าหรือเท่ากับ 280

	กล้ามเนื้อหัวใจวาย	
	มี	ไม่มี
CPK ≥ 280	97	1
CPK < 280	133	129

Sensitivity = $\frac{97}{97+133} \times 100 = 42.2\%$

Specificity = $\frac{129}{1+129} \times 100 = 99.2\%$

เมื่อใช้จุดตัดของ CPK มากกว่าหรือเท่ากับ 80

		กล้ามเนื้อหัวใจวาย	
		มี	ไม่มี
CPK ≥ 80		215	16
CPK < 80		15	114
Sensitivity	=	$\frac{215}{215+15} \times 100 = 93.5\%$	
Specificity	=	$\frac{114}{16+114} \times 100 = 87.7\%$	

เมื่อใช้จุดตัดของ CPK มากกว่าหรือเท่ากับ 40

		กล้ามเนื้อหัวใจวาย	
		มี	ไม่มี
CPK ≥ 40		228	42
CPK < 40		2	88
Sensitivity	=	$\frac{228}{228+2} \times 100 = 99.1\%$	
Specificity	=	$\frac{88}{42+88} \times 100 = 67.7\%$	

เมื่อใช้จุดตัดของ CPK มากกว่าหรือเท่ากับ 0

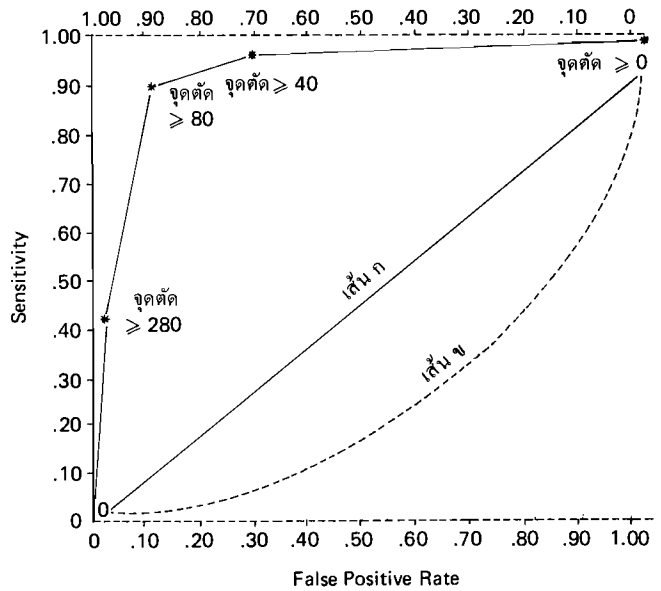
		กล้ามเนื้อหัวใจวาย	
		มี	ไม่มี
CPK ≥ 0		230	130
CPK < 0		0	0
Sensitivity	=	$\frac{230}{230+0} \times 100 = 100\%$	
Specificity	=	$\frac{0}{130+0} \times 100 = 0\%$	

จากผลการศึกษาดังกล่าวสามารถหาค่า sensitivity, specificity และค่า false positive value = 100 - specificity (%) ของจุดตัดแต่ละจุดได้ดังในตารางที่ 2

7.2 การสร้าง ROC curve จากตารางข้างบนนี้จะนำค่า CPK ที่เป็นจุดตัดแต่ละจุดพร้อมค่าของ sensitivity, specificity และ false positive value ของแต่ละจุดตัดมาสร้างเป็น receiver operating characteristic (ROC) curve ได้ดังรูป จะพบว่าจุดตัดที่ 80 จะอยู่ใกล้มุมซ้ายบนมากที่สุด คือ ใกล้ค่า 1 ของ sensitivity และ specificity มากที่สุด จุดตัดนี้จึงเหมาะสมที่สุดในกรณีที่ต้องการได้ค่าบวกเท็จและผลลบเท็จจนน้อยที่สุด นั่นคือผลรวมของ sensitivity และ specificity จะมีค่าสูงสุดที่จุดตัดนี้

ถ้าลากได้เส้นตรงหรือเกือบเป็นเส้นตรง (เส้น ก) แสดงว่าวิธีทดสอบมีคุณค่าน้อยมาก หรือมีคุณค่าเท่ากับการจับฉลากให้เป็นผลบวกหรือลบ แต่ถ้าได้เส้นโค้งเว้าออกห่างจากค่า 1 ของ sensitivity และ specificity (เส้น ข) แสดงว่ามีการกำหนดนิยามหรือคำนวณผิดไปทั้งระบบ (systemic error) ในกรณีสุดท้ายนี้ให้ปรับเปลี่ยนแล้วจะพบว่าวิธีทดสอบนั้นก็มีประโยชน์เช่นกัน

7.3 การใช้จุดตัดหลายจุดเพื่อตรวจคุณภาพของวิธีทดสอบ เมื่อแบ่งผลการทดสอบออกเป็นหลายๆ ช่วง โดยใช้จุดตัดหลายจุด



รูปที่ 1

ตารางที่ 2

จุดตัดที่กำหนดค่า CPK มากกว่าหรือเท่ากับให้เป็นผลบวก	Sensitivity (%)	Specificity (%)	False positive value (%)
280	42	99	1
80	93	88	12
40	99	68	32
0	100	0	100

เราจะสามารถคำนวณค่า sensitivity และ specificity ที่จุดตัดแต่ละจุดได้ การเลือกจุดตัดที่เหมาะสมต้องพิจารณาหลายด้าน แต่ถ้าต้องการจุดตัดที่จะให้ผลรวมของ sensitivity และ specificity มีค่าสูงสุดโดยมีผลบวกเท็จและผลลบเท็จมีค่าน้อยที่สุด เราสามารถเลือกจุดตัดตามที่ต้องการได้โดยนำค่า sensitivity และ specificity มารวมกัน ผลรวมของคูใดให้ค่าสูงสุดเราก็จะเลือกจุดตัดที่ให้ค่า sensitivity และ specificity คู่่นั้นเป็นจุดตัดที่เหมาะสมดังที่ได้อธิบายมาแล้ว

นอกจากนี้เรายังสามารถพิจารณาประโยชน์ของวิธีทดสอบดังกล่าวได้จากผลรวมของ sensitivity และ specificity ที่เกิดจากจุดตัดแต่ละจุด ผลรวมของ sensitivity และ specificity ของทุกคู่อิงมีค่าสูงเกิน 150 และเข้าใกล้ 200 มากเท่าใด วิธีทดสอบนั้นจะยิ่งมีประโยชน์มากขึ้น แต่ถ้าผลรวมของแต่ละคู่อิงมีค่าใกล้ 100 วิธีทดสอบนั้นเกือบจะไม่มีประโยชน์เลย เพราะการโยนเหรียญออกหัว-ก้อยก็จะได้ค่าใกล้เคียง 100 เช่นกัน และยังไม่ต้องเสียเวลาและทุนทรัพย์ในการทำการทดสอบด้วย

ถ้าผลรวมของแต่ละคู่อิงมีค่าต่ำกว่า 100 แสดงว่าต้องมีความ

ตารางที่ 3 แสดงผลรวมของ sensitivity และ specificity ของวิธีทดสอบที่มีประโยชน์ในการวินิจฉัยโรค

จุดตัดที่	Sensitivity (%)	Specificity (%)	ผลรวม
ก	94	72	166
ข	88	85	173
ค	82	88	170
ง	78	92	170
จ	65	95	160

จะพบว่าจุดตัด ข ให้ผลรวมสูงสุด และผลรวมที่จุดตัดแต่ละจุดมีค่าใกล้ 200 แสดงว่าวิธีทดสอบนี้มีคุณค่าในการวินิจฉัยโรค ถ้าจะให้ดีที่สุดจะต้องมีผลรวมเท่ากับ 200 อย่างน้อย 1 ครั้ง ถ้าได้ 200 หลายครั้งที่จุดตัดหลายจุดจะเป็นวิธีทดสอบที่มีคุณค่ามากที่สุด

ตารางที่ 4 แสดงผลรวมของ sensitivity และ specificity ของวิธีทดสอบที่ไม่มีประโยชน์ในการวินิจฉัยโรค

จุดตัดที่	Sensitivity (%)	Specificity (%)	ผลรวม
ก	65	36	101
ข	60	40	100
ค	55	45	100
ง	50	52	102
จ	45	57	102

จะพบว่าผลรวมที่จุดตัดแต่ละจุดใกล้เคียง 100 แสดงว่าวิธีทดสอบนี้ไม่มีคุณค่าในการวินิจฉัยโรค

ผิดพลาดชนิด systematic error เช่น การกำหนดเครื่องหมายผิดสำหรับทุกค่าที่วัดได้ หรือการให้ค่าจำกัดความผิดหรือตรงกันข้ามกับที่ควรเป็น ในกรณีนี้ให้กลับมาพิจารณาค่าจำกัดความหรือกระบวนการทดสอบการคำนวณหรือการใส่เครื่องหมายใหม่ ก็จะสามารถแก้ไขและทำให้วิธีทดสอบกลับมามีประโยชน์ได้

การใช้ผลรวมของ sensitivity และ specificity ที่จุดตัดแต่ละจุดเพื่อพิจารณาเลือกจุดตัดและความมีประโยชน์ของวิธีทดสอบจะแสดงในตัวอย่างดังตารางที่ 3 และ 4

8. การใช้จุดตัดหลายจุดและการคำนวณหาค่า likelihood ratio

เนื่องจากความเหลื่อมล้ำของค่าตัวแปร (ค่าที่วัดได้จากการทดสอบ) ระหว่างประชากรศึกษาแต่ละกลุ่มเป็นสิ่งที่หลีกเลี่ยงยาก การกำหนดจุดตัดเพียง 1 จุดจึงทำให้เกิดผลบวกและลบเท่านั้น และยังมีผลบวกเท็จและผลลบเท็จด้วยเสมอ การมีจุดตัดเพียง 1 จุดสำหรับแบ่งประชากรที่เป็นโรคออกจากกลุ่มอื่น จึงไม่สามารถแสดงความน่าจะเป็นโรคในผู้ป่วยแต่ละรายว่ามีมากน้อยเท่าใด กล่าวคือไม่สามารถแสดงความน่าจะเป็นโรคว่ามีมากน้อยเพียงใดในประชากรแต่ละรายที่ให้ผลบวกหรือผลลบแต่มีค่าแตกต่างกัน ผู้ป่วยบางรายให้ค่าตัวแปรสูงมาก ซึ่งอาจจะเป็นค่าที่สูงชัดเจน เป็นไปตามแบบฉบับของโรค ดังนั้นโดยทฤษฎีแล้วค่าที่วัดได้สูงมากก็น่าจะเป็นโรคนั้นมาก ถ้าวัดได้ค่าน้อยก็ไม่น่าจะเป็นโรค การกำหนดจุดตัดเพียงจุดเดียวจึงเกิดจุดอ่อนขึ้นเพราะไม่สามารถแสดงความน่าจะเป็นโรคได้ชัดเจนตามผลการทดสอบ วิธีแก้ไขจุดอ่อนดังกล่าวนี้กระทำได้โดยแบ่งค่าตัวแปรที่วัดได้ออกเป็นช่วงต่างๆ แล้วแสดง *ความเป็นเท่า* ของความน่าจะเป็นโรคในค่าตัวแปรแต่ละช่วงออกมา *ความเป็นเท่า* ของความน่าจะเป็นโรสดังกล่าวมีชื่อว่า *likelihood ratio* ซึ่งได้มาจากการคำนวณดังในตารางที่แสดงวิธีการคำนวณค่า likelihood ratio ของค่าตัวแปรแต่ละช่วงที่วัดได้จากการทดสอบหา CPK ซึ่งใช้ผลการศึกษาในหัวข้อที่ 7 มาแสดงวิธีคำนวณ (ตารางที่ 5)

จากตารางดังกล่าวจะได้ค่า likelihood ratio เท่ากับ 55 สำหรับผู้ที่ได้ค่า CPK เท่ากับหรือเกิน 280 หน่วย แสดงว่าผู้ป่วยรายหนึ่งที่มีค่า CPK ในเลือดเท่ากับ 320 หน่วยจะมีโอกาสเป็นโรคกล้ามเนื้อหัวใจวายถึง 55 เท่า ค่า likelihood ratio จะเป็นค่าคงที่และไม่แปรปรวนไปตามความชุกของโรค ข้อดีอีกประการหนึ่งคือสามารถใช้ likelihood ratio ในการคำนวณความชุกของโรค (หรือความน่าจะเป็นโรคนั้น) หลังจากทราบผลการทดสอบได้ด้วยตามวิธีคำนวณที่จะแสดงต่อไป

8.1 ขั้นตอนการคำนวณเพื่อแสดงวิธีประยุกต์ใช้ likelihood ratio และศัพท์ที่ใช้ในการคำนวณ วิธีคำนวณมี 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 เปลี่ยนความชุกเป็น pretest odds ก่อนโดยใช้สูตร

$$\text{pretest odds} = \frac{\text{probability}}{1 - \text{probability}}$$

ตารางที่ 5 แสดงการคำนวณ likelihood ratio สำหรับค่า CPK ในช่วงต่างๆ

		โรคกล้ามเนื้อหัวใจตาย				likelihood ratio
		มีโรค		ไม่มีโรค		
		จำนวน	proportion	จำนวน	proportion	
ค่า CPK	> 280	97	$\frac{97}{230} = 0.4217$	1	$\frac{1}{130} = 0.0077$	$\frac{0.4217}{0.0077} = 55$
	80 – 279	118	$\frac{118}{230} = 0.5131$	15	$\frac{15}{130} = 0.1154$	$\frac{0.5131}{0.1154} = 4.4$
	40 – 79	13	$\frac{13}{230} = 0.0565$	26	$\frac{26}{130} = 0.2000$	$\frac{0.0565}{0.2000} = 0.3$
	1 – 39	2	$\frac{2}{230} = 0.0087$	88	$\frac{88}{130} = 0.6769$	$\frac{0.0087}{0.6769} = 0.01$
รวม		230	1.0	130	1.0	

ขั้นตอนที่ 2 คุณ pretest odds ด้วย likelihood ratio ผลลัพธ์ที่ได้จะเป็น post-test odds

ขั้นตอนที่ 3 เปลี่ยน post-test odds เป็น probability หลังทำการทดสอบ โดยใช้สูตร

$$\text{Post-test probability} = \frac{1}{1 + \text{post-test odds}}$$

8.2 ตัวอย่างแสดงวิธีประยุกต์ใช้ likelihood ratio ทางคลินิก การประยุกต์ใช้ likelihood ratio จะต้องทราบค่าความชุกของโรค ค่าตัวแปรที่วัดได้จากการทดสอบในผู้ป่วยแต่ละราย และค่า likelihood ratio ซึ่งมีค่าคงที่และเป็นตัวเลขประจำค่าตัวแปรแต่ละช่วง ดังตัวอย่างที่จะแสดงต่อไปนี้ (ค่าต่างๆ เป็นค่าที่สมมติขึ้น)

ความชุกของกล้ามเนื้อหัวใจตายก่อนที่จะรับย้ายเข้าหออภิบาลโรคหัวใจ (ที่ตึกตรวจโรคนอก) = 0.3 (หรือร้อยละ 30)

แพทย์รับผู้ป่วย ก เข้าหออภิบาลโรคหัวใจและตรวจเลือดได้ค่า CPK = 340 หน่วย

แพทย์รับผู้ป่วย ข เข้าหออภิบาลโรคหัวใจและตรวจเลือดได้ค่า CPK = 26 หน่วย

ให้ดูค่า likelihood ratio ของค่า CPK แต่ละช่วง ในตารางแสดงการคำนวณค่า likelihood ratio

8.3 วิธีการคำนวณ post-test probability เมื่อใช้วิธีทดสอบ 1 วิธี

ขั้นที่ 1 เปลี่ยนความชุกเป็น pre-test odds ของผู้ป่วย ก และ ข ซึ่งเป็นค่าก่อนที่จะรับเข้าหออภิบาลโรคหัวใจ (ค่าความชุกของโรคของผู้ป่วย ก และ ข มีค่าเท่ากัน เพราะมีลักษณะคลินิกเหมือนกัน)

$$\text{pre-test odds} = \frac{0.3}{1-0.3} = \frac{0.3}{0.7} = 0.4286$$

ขั้นที่ 2 คุณ pre-test odds ด้วย likelihood ratio

(LR) ของค่า CPK ที่ได้จากผู้ป่วยแต่ละราย เพื่อคำนวณหา post-test odd ของผู้ป่วย ก และ ข

ผู้ป่วย ก (CPK = 430, LR = 55), $55 \times 0.4286 = 23.573$

ผู้ป่วย ข (CPK = 26, LR = .01), $0.01 \times 0.4286 = 0.004286$

ขั้นที่ 3 เปลี่ยน post-test odds เป็น probability ตามผลการทดสอบตัวอย่างของผู้ป่วยแต่ละราย

$$\text{ผู้ป่วย ก post-test probability} = \frac{23.573}{1 + 23.573} = 0.96$$

$$\text{ผู้ป่วย ข post-test probability} = \frac{0.004286}{1 + 0.004286} = 0.004$$

ผลการคำนวณดังกล่าวจะสรุปว่า ผู้ป่วย ก มีความน่าจะเป็นโรคกล้ามเนื้อหัวใจตายสูงขึ้นเป็นร้อยละ 96 เมื่อตรวจพบ CPK มีค่าเท่ากับ 340 หน่วย ส่วนผู้ป่วย ข มีความน่าจะเป็นโรคเหลือเพียงร้อยละ 0.4 เท่านั้น เมื่อตรวจพบ CPK มีค่าเท่ากับ 26 หน่วย นั่นคือผู้ป่วย ข ไม่น่าจะเป็นโรคนี้

โดยสรุปประโยชน์ของ likelihood ratio มีดังนี้

1. สามารถนำมาใช้คำนวณหา post-test probability ของโรคได้รวดเร็วเมื่อทราบผลการทดสอบ

2. สามารถจำแนกน้ำหนักความน่าจะเป็นโรคจากผลการทดสอบได้ นั่นคือ การทดสอบยิ่งให้ผลบวกมาก ความน่าจะเป็นโรคจะเพิ่มมากขึ้นด้วย ทำนองเดียวกันถ้าได้ผลลบมากเท่าใดก็ไม่น่าจะเป็นโรคมากเท่านั้น

3. ในกรณีที่ใช้การทดสอบมากกว่า 1 วิธีเพื่อวินิจฉัยโรค 1 โรค และทราบค่า likelihood ratio สำหรับการทดสอบแต่ละช่วงที่ได้จากการทดสอบแต่ละวิธี จะสามารถคำนวณหาความน่าจะเป็นโรคของผู้ป่วยที่ได้รับการทดสอบหลายๆวิธีได้อย่างรวดเร็ว ดังที่จะแสดงต่อไป

8.4 วิธีการคำนวณ post-test probability เมื่อใช้วิธีทดสอบมากกว่า 1 วิธี ในกรณีที่มีการวินิจฉัยโรคอาศัยผลการทดสอบมากกว่า 1 อย่าง การคำนวณหาความน่าจะเป็นโรคเมื่อทราบผลการทดสอบทั้งหมด จะปฏิบัติได้ง่ายโดยถ้าใช้ likelihood ratio ในการคำนวณโดยมีขั้นตอนดังนี้

8.4.1 เปลี่ยน pre-test probability เป็น pre-test odds

8.4.2 คูณ pre-test odds ด้วย likelihood ratio ของผลการทดสอบทุกชนิด เมื่อคูณจนหมดแล้วจะได้ค่าสุดท้ายเป็น post-test odds

8.4.3 เปลี่ยน post-test odds เป็น post-test probability

จะพบว่าหัวข้อที่ 8.4.1 และ 8.4.3 ก็เหมือนกับขั้นตอนที่ 1 และ 3 ในหัวข้อที่ 8.1 นั่นเอง กล่าวคือ การเปลี่ยน pre-test probability เป็น pre-test odds และเปลี่ยนจาก post-test odds มาเป็น post-test probability ยังใช้วิธีการคำนวณเหมือนเดิม ส่วนหัวข้อที่ 8.4.2 จะเขียนเป็นสูตรได้ดังนี้

Pre-test odds × LR (วิธีทดสอบ 1) × LR (วิธีที่ 2) × LR (วิธีทดสอบที่ 3) = post-test odds

ตัวอย่างสมมติที่ใช้เพื่อแสดงการคำนวณ post-test probability ของฝีในตับจากเชื้อทรงแท่งแกรมลบ โดยใช้ผลการทดสอบ 2 วิธี เนื่องจากฝีในตับมีสาเหตุจากเชื้อต่างๆ เช่น เชื้อบิด, เชื้อแบคทีเรียทรงกลมแกรมบวก, เชื้อแบคทีเรียทรงแท่งแกรมลบชนิดฟังหรือไม่ฟังออกซิเจน ฯลฯ การเลือกใช้ยาต้านจุลชีพสำหรับเชื้อแต่ละกลุ่มมีความแตกต่างกันและบางครั้งไม่สามารถเก็บตัวอย่างมาตรวจได้ ผู้วิจัยจึงพิจารณาใช้ตัวแปร 2 ชนิดมาวินิจฉัยเชื้อทรงแท่งแกรมลบ เพราะเป็นโรคที่ต้องใช้ยาแพงและรักษาติดต่อกันนาน สมมติว่าการทดสอบ 2 วิธีที่เลือกใช้ในการทำนายเชื้อมีดังนี้

การทดสอบวิธีที่ 1 คือ จำนวนเม็ดเลือดขาว/ลบ.มม.

การทดสอบวิธีที่ 2 คือ ระดับ bilirubin/ดล.

เมื่อพบผู้ป่วยที่มีไข้และแพทย์ตรวจพบแหล่งติดเชื้อภายในร่างกายคือฝีในตับ แต่ไม่สามารถเก็บตัวอย่างมาตรวจชนิดของเชื้อแบคทีเรียได้ในขณะนั้น จึงได้ค้นคว้าและพบผลการศึกษาในอดีตเกี่ยวกับจำนวนเม็ดเลือดขาวและระดับ bilirubin ในการทำนายเชื้อทรงแท่งแกรมลบของฝีในตับได้ผลดังนี้

ความสัมพันธ์ของ likelihood ratios และจำนวนเม็ดเลือดขาวในแต่ละช่วง

จำนวนเม็ดเลือดขาว (เซลล์/ลบ.มม.)	likelihood ratios
ต่ำกว่า 5,000	9.0
5,000 - 10,000	6.2
10,000 - 20,000	2.1
มากกว่า 20,000	0.01

ความสัมพันธ์ของ likelihood ratios และระดับ bilirubin ในแต่ละช่วง

ระดับ bilirubin (มก./ดล.)	likelihood ratios
0.1-1.0	0.1
1.1 - 2.0	0.3
2.0 - 4.0	1.6
4.0 - 8.00	9.0
มากกว่า 8.00	17.0

ผู้ป่วย ค มีไข้ เจ็บที่ชายโครงขวา มารับการรักษาที่สถานพยาบาลแห่งหนึ่งที่มีผู้สำรวจในปีที่ผ่านมา พบว่าความชุกของฝีในตับจากเชื้อทรงแท่งแกรมลบมีค่าเท่ากับร้อยละ 1 แพทย์ตรวจผู้ป่วยแล้วสงสัยว่าเป็นฝีในตับ การตรวจเลือดได้ค่าเม็ดเลือดขาว 8,250 เซลล์/ลบ.มม. และ bilirubin มีค่าเท่ากับ 5.2 มก./ดล. อยากทราบว่าเมื่อทราบผลการทดสอบดังกล่าวแล้วจะทำให้แพทย์ทำนายว่าผู้ป่วย ค จะมีความน่าจะเป็นโรคฝีในตับจากเชื้อทรงแท่งแกรมลบมากเพียงใด?

8.5 วิธีการคำนวณเมื่อใช้วิธีทดสอบ 2 วิธี

ขั้นที่ 1 เปลี่ยนความชุกเป็น pre-test odds ของผู้ป่วย ค ซึ่งเป็นค่าก่อนจะรับการทดสอบ (ความชุกของโรคเท่ากับร้อยละ 1 หรือ probability = 0.01)

$$\text{pre-test odds} = \frac{0.01}{1 - 0.01} = 0.0101$$

ขั้นที่ 2 คูณค่า pre-test odds ด้วย likelihood ratio (LR) ของค่าจำนวนเม็ดเลือดขาวและระดับ bilirubin ของผู้ป่วย ค เพื่อคำนวณหา post-test odds ของผู้ป่วย ค

$$\text{LR ของเม็ดเลือดขาว 8,250 เซลล์/ลบ.มม.} = 6.2$$

$$\text{LR ของ bilirubin 5.2 มก./ดล.} = 9.0$$

$$\text{post-test odd} = 0.0101 \times 6.2 \times 9.0 = 0.56358$$

ขั้นที่ 3 เปลี่ยน post-test odds เป็น probability ของผู้ป่วย ค ตามผลการทดสอบตัวอย่างทั้ง 2 วิธี

$$\text{post-test probability} = \frac{0.56358}{1 + 0.56358} = 0.36044$$

ผลการคำนวณจะสรุปว่า โอกาสที่เชื้อก่อโรคจะเป็นเชื้อทรงแท่งแกรมลบ มีค่าเท่ากับ 0.36044 หรือร้อยละ 36 จากผลการทดสอบทั้ง 2 วิธีซึ่งแพทย์ผู้รักษาจะต้องตัดสินใจต่อไปว่าจะเลือกใช้ยาต้านจุลชีพที่มุ่งทำลายเชื้อ *E. coli* หรือ *Klebsiella* หรือไม่อย่างไรก็ตามผลการทดสอบทั้ง 2 วิธีทำให้ความน่าจะเป็นโรคจากเชื้อทรงแท่งแกรมลบจากร้อยละ 1 มาเป็นร้อยละ 36

8.6 การคำนวณ likelihood ratio จากค่า sensitivity และ specificity ในกรณีที่ทราบค่า sensitivity และ specificity ของวิธีทดสอบนั้นแล้ว เราสามารถคำนวณค่า likelihood ratio ได้เช่นกัน โดยเปลี่ยนค่าร้อยละของทั้ง 2 ค่าให้เป็น probability ก่อน แล้วใช้สูตรดังนี้

$$\text{Likelihood ratio} = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

ยกตัวอย่าง เช่น การทดสอบ ก มีค่า sensitivity = ร้อยละ 70 และมีค่า specificity = ร้อยละ 95 ณ จุดตัดจุดหนึ่งของผลการทดสอบ การคำนวณหา likelihood ratio กระทำโดย

$$\begin{aligned} \text{Likelihood ratio} &= \frac{\text{sensitivity}}{1 - \text{specificity}} \\ &= \frac{0.70}{1 - 0.95} \\ &= \frac{0.7}{0.05} \\ \text{Likelihood ratio} &= 14 \end{aligned}$$

ถ้าสังเกตดูจะพบว่า สูตรคำนวณดังกล่าวก็เป็นวิธีแบบเดียวกับการคำนวณหา likelihood ratio จากข้อมูลดิบเช่นกัน เนื่องจากรูปทดสอบต่างๆ ในปัจจุบันใช้จุดตัดเพียงจุดเดียว ค่า sensitivity และ specificity จึงมีเพียง 1 ค่า การคำนวณหา likelihood ratio ในกรณีนี้จึงได้เพียง 2 ค่าเท่านั้น คือ ได้ค่า likelihood ratio

ของผลบวกและผลลบเท่านั้น แต่ถ้าวิธีทดสอบนั้นแสดงจุดตัดหลายๆ จุด ซึ่งมีผลให้ค่า sensitivity และ specificity เปลี่ยนตามจุดตัดแต่ละจุดไปด้วย เราจะสามารถคำนวณหา likelihood ratio ของจุดตัดแต่ละจุดได้และทำให้ผลการทดสอบที่ได้มี "น้ำหนักของความน่าจะเป็นหรือไม่เป็นโรค" ได้ด้วย ซึ่งจะนำไปตามเป้าหมายอย่างหนึ่งของการศึกษาเรื่อง likelihood ratio ตัวอย่างของ likelihood ratio ที่มีผู้ศึกษาไว้แล้วแสดงดังตารางที่ 6

8.7 ตารางสำเร็จรูปและ nomogram ในการประยุกต์ใช้ likelihood ratio เนื่องจาก likelihood ratio มีประโยชน์ดังกล่าวแล้วในการแปลผลการทดสอบและสามารถนำมาใช้คำนวณหาความน่าจะเป็นโรคออกมาเป็นตัวเลขได้ แต่การคำนวณเปลี่ยน pre-test probability เป็น pre-test odds, การคูณ pre-test odds กับ likelihood ratio และเปลี่ยน post-test odds มาเป็น post-test

ตารางที่ 6 แสดงค่า likelihood ratios ของค่าต่างๆ ที่ได้จากวิธีทดสอบตามที่เคยมีการศึกษาไว้แล้ว

Diagnosis test	Target disorder*	Result	Likelihood ratio
Symptoms of typical angina	Coronary artery stenosis (by angiography or autopsy)	Positive history	
		Men	115
		Women	120
Symptoms of atypical angina	Coronary artery stenosis (by angiography or autopsy)	Positive history	
		Men	14
		Women	15
Exercise electrocardiography	Coronary artery stenosis (by angiography)	Non-sloping ST segment depression of:	
		≥ 2.5 mm.	39
		2 - 2.49	11
		1.5 - 1.99	4.2
		1 - 1.49	2.1
		.05 - .99	.92
		< .05	.23
Ultrasound		Definitely abnormal	5.6
		Probably abnormal	2.1
		Possibly abnormal	.95
		Probably normal	.43
		Definitely normal	.32
CT scan	Pancreatic Disease (by biopsy, autopsy or clinical course)	Definitely abnormal	26
		Probably abnormal	4.8
		Possibly abnormal	.35
		Probably normal	.32
		Definitely normal	.11
Carcinoembryonic antigen	Colorectal cancer (by biopsy or operation)	≥ 20 ng/ml	3.5
		10 - 19.9	2.3
		5 - 9.9	1.4
		1 - 4.9	.94
		< 1 ng/ml	.46

*วิธีทดสอบมาตรฐาน (gold standard) อยู่ในวงเล็บ

ตารางที่ 7 ตารางสำเร็จรูปในการเปลี่ยน pre-test probabilities ให้เป็น pre-test odds

คิดเป็นร้อยละ	หรือ	ความน่าจะเป็นโรค		
		pre-test probabilities	probability 1-probability	pre-test odds
.1	หรือ	.001	.001/.999	.001
1	หรือ	.01	.01/.99	.01
2	หรือ	.02	.02/.98	.02
3	หรือ	.03	.03/.97	.03
4	หรือ	.04	.04/.96	.04
5	หรือ	.05	.05/.95	.05
10	หรือ	.1	.1 / .9	.11
20	หรือ	.2	.2 / .8	.25
30	หรือ	.3	.3 / .7	.43
40	หรือ	.4	.4 / .6	.67
50	หรือ	.50	.5 / .5	1.0
60	หรือ	.6	.6 / .4	1.5
70	หรือ	.7	.7 / .3	2.3
80	หรือ	.8	.8 / .2	4.0
90	หรือ	.9	.9 / .1	9.0
95	หรือ	.95	.95 / .05	19.0
99	หรือ	.99	.99 / .01	99.0

ที่จริงแล้ว ค่า pre-test odd ที่คำนวณได้จะเป็นอัตราส่วนต่อ 1 เสมอ เช่น กำหนด pre-test odd ได้ 0.001 หมายถึงว่า 0.001:1 เสมอ แต่เพื่อความคล่องตัวในการปฏิบัติงานจึงได้ละไว้ในฐานที่เข้าใจ

probability อีก ทำให้ต้องเสียเวลาคำนวณ จึงมีผู้สร้างตารางสำเร็จรูปในการเปลี่ยน pre-test probabilities เป็น pre-test odds ดังในตารางที่ 7, เปลี่ยน post-test odds เป็น post-test probability ให้รวดเร็วดังในตารางที่ 8 หรือเปลี่ยนจาก pre-test probability เป็น post-test probability โดยการลากเส้นผ่านค่า likelihood ratio ที่ได้จากการทดสอบดังที่จะแสดงในรูปที่ 2

ตัวอย่างการใช้ nomogram

สมมติว่า ความชุกของโรคกล้ามเนื้อหัวใจวายที่ตีตรวจโรค นอก ของโรงพยาบาลแห่งหนึ่งมีค่าเท่ากับร้อยละ 1 ผู้ป่วยรายหนึ่งมีอาการเจ็บแน่นหน้าอกมารับการเจาะเลือดหา CPK ได้ค่า 302 หน่วย เมื่อเปิดตารางค่า CPK พบว่าค่า CPK ที่ได้เกิน 280 หน่วย จะมีค่า likelihood ratio เท่ากับ 55 วิธีใช้ nomogram สำหรับผู้ป่วยรายนี้มีดังนี้

Pre-test probability (percent probability) = 1%
Likelihood ratio สำหรับค่า CPK เท่ากับ 302 หน่วย = 55

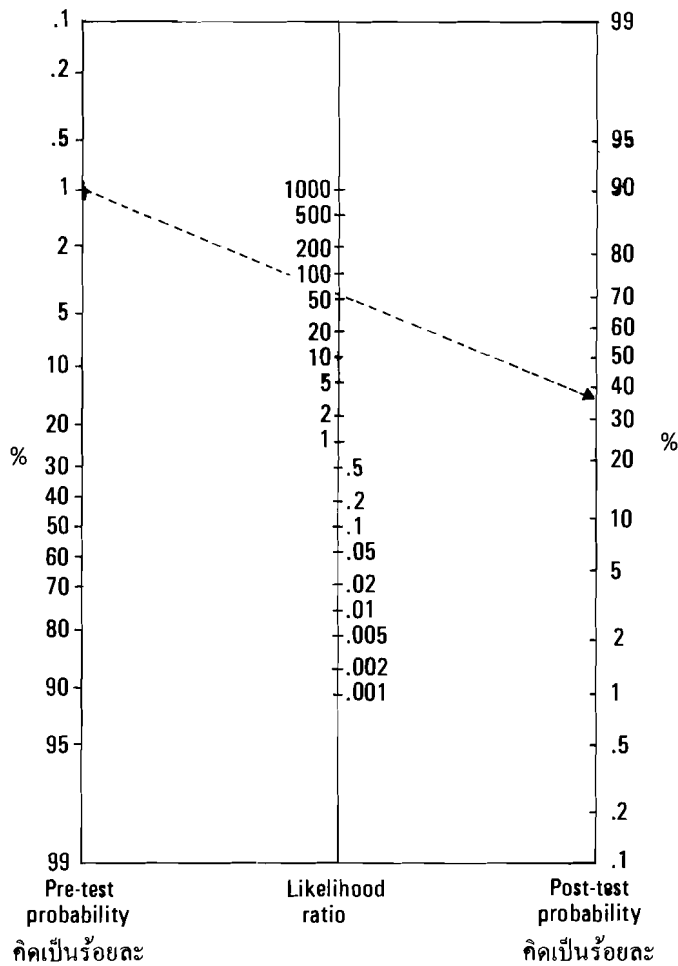
ตารางที่ 8 ตารางสำเร็จรูปในการเปลี่ยน post-test odds ให้เป็น probabilities หรือร้อยละหลังทราบผลการทดสอบ

Post-test odds	หรือ	ความน่าจะเป็นโรค			หรือ	คิดเป็นร้อยละ
		odds	odds + 1	post-test probabilities		
.001	หรือ	.001/1.001		.001	หรือ	.1
.01	หรือ	.01 /1.01		.01	หรือ	1
.02	หรือ	.02 /1.02		.02	หรือ	2
.03	หรือ	.03 /1.03		.03	หรือ	3
.04	หรือ	.04 /1.04		.04	หรือ	4
.05	หรือ	.05 /1.05		.05	หรือ	5
.1	หรือ	.1 /1.1		.09	หรือ	9
.2	หรือ	.2 /1.2		.17	หรือ	17
.3	หรือ	.3 /1.3		.23	หรือ	23
.4	หรือ	.4 /1.4		.29	หรือ	29
.5	หรือ	.5 /1.5		.33	หรือ	33
.6	หรือ	.6 /1.6		.38	หรือ	38
.7	หรือ	.7 /1.7		.41	หรือ	41
.8	หรือ	.8 /1.8		.44	หรือ	44
.9	หรือ	.9 /1.9		.47	หรือ	47
1.9	หรือ	1/2		.5	หรือ	50
2.0	หรือ	2/3		.67	หรือ	67
3.0	หรือ	3/4		.75	หรือ	75
4.0	หรือ	4/5		.8	หรือ	80
5.0	หรือ	5/6		.83	หรือ	83
10.0	หรือ	10/11		.91	หรือ	91
20.0	หรือ	20/21		.95	หรือ	95
30.0	หรือ	30/31		.97	หรือ	97
40.0	หรือ	40/41		.98	หรือ	98
50.0	หรือ	50/51		.98	หรือ	98
100.0	หรือ	100/101		.99	หรือ	99

ลากเส้นตรงจาก pre-test probability ตรง 1% ผ่านเส้นกลาง (likelihood ratio) ที่ค่า 55 จะได้เส้นตรงต่อมาชนเส้น post-test probability คิดเป็น percent probability ที่ค่า 35% นั่นคือ ความน่าจะเป็นโรคกล้ามเนื้อหัวใจวายที่ตีตรวจโรค นอก (pre-test probability) มีค่าเท่ากับร้อยละ 1 แต่เมื่อเจาะเลือดได้ค่า CPK ดังกล่าว ความน่าจะเป็นโรคกล้ามเนื้อหัวใจวายสูงขึ้นมาเป็นร้อยละ 35 (post-test probability)

9. จุดอ่อนของ likelihood ratio

อย่างไรก็ตาม ถึงแม้ likelihood ratio จะเป็นวิธีที่เหมาะสม



รูปที่ 2 Nomogram ในการประยุกต์ใช้ likelihood ratio เพื่อปรับความน่าจะเป็นโรคหลังทราบผลการทดสอบ

มากขึ้นในการให้นำหนักความน่าจะเป็นโรคของค่าผลการทดสอบที่ให้ผลบวกหรือผลลบก็ตาม แต่เนื่องจากตัวแปร (ผลการทดสอบ) ที่วัดได้แต่ละตัวจากการทดสอบมักจะไม่เป็นอิสระต่อกัน การใช้ likelihood ratio ของตัวแปรที่ไม่เป็นอิสระต่อกัน จะทำให้ post-test probability หลังการทดสอบหลายๆ ชนิดได้ค่าสูงกว่าที่เป็นจริงหรือต่ำกว่าที่เป็นจริง ยกตัวอย่าง เช่น การตรวจเลือดผู้ป่วยที่มีดีซ่านรายหนึ่งได้ค่า SGOT และ SGPT เท่ากับ 980 และ 1,020 หน่วยตามลำดับ ในกรณีเช่นนี้ถ้า SGPT มีค่าสูงขึ้น SGOT มักจะสูงตามด้วย ค่าของ SGOT จึงไม่เป็นอิสระจากค่าของ SGPT ในทางสถิติจะกลายเป็นว่า SGOT เป็น dependent variable สมมติว่าค่า likelihood ratio ของ SGPT ที่มีค่าเท่ากับ 1,020 มีค่าเท่ากับ 78 และ likelihood ratio ของ SGOT ที่มีค่าเท่ากับ 980 มีค่าเท่ากับ 72 ถ้านำค่า likelihood ratio ทั้ง 2 ค่านี้มาคูณกับ pre-test odds กลายเป็น post-test odds หลังผลการทดสอบ 2 ชนิด แล้วคำนวณกลับเป็น post-test probability จะได้ค่าสูง

กว่าที่เป็นจริงมาก ในกรณีเช่นนี้ควรใช้ likelihood ratio เพียงค่าเดียว เช่น SGPT มาคำนวณเท่านั้น ถ้ามีการทดสอบวิธีอื่นที่ให้ผลเป็นอิสระ (independent variable) ต่อค่า SGPT จึงจะนำค่า likelihood ratio มาคำนวณได้

จุดอ่อนอีกประการหนึ่งของ likelihood ratio คือ ผลการทดสอบหลายๆ ชนิดที่วัดได้มิได้มีไว้สำหรับผลลัพธ์เพียงอย่างหนึ่งและการที่จะนำผลการทดสอบหลายๆ ชนิดมาคำนวณโดยใช้ likelihood ratio จะมีปัญหาหลายประการ และทำให้ post-test probability ของโรคหลังจากคำนวณเสร็จแล้วอาจจะคลาดเคลื่อนไปจากค่าที่เป็นจริงได้มาก เพราะค่า likelihood ratio ที่ได้จากการทดสอบต่างๆ มักจะไม่ได้มาจากการศึกษาครั้งเดียวกัน แต่จะศึกษามาจากกลุ่มประชากรที่มีลักษณะแตกต่างกันมากหลายกลุ่ม นอกจากนี้ค่าตัวแปรต่างๆ ที่วัดได้จากผลการทดสอบหลายๆ ชนิดต่อผลลัพธ์อย่างเดียวกัน จะมีทั้งตัวแปรที่ต่อเนื่อง (continuous) และไม่ต่อเนื่อง (discrete) และเมื่อนำมารวมกันเพื่อวินิจฉัยโรคใดโรคหนึ่ง นำหนักของตัวแปรเหล่านี้ที่นำมารวมกันเพื่อวินิจฉัยโรคนั้นอาจจะไม่เท่ากัน จึงต้องมีการแก้ไข "น้ำหนัก" ของค่าตัวแปรแต่ละอย่างให้เกิดประโยชน์สูงสุดในการวินิจฉัยแยกโรคชนิดนั้นออกจากโรคต่างๆ นอกจากนี้น้ำหนักของค่าตัวแปรต่างๆ อาจจะเปลี่ยนแปลงไปอีก ถ้านำค่าตัวแปรนั้นไปวินิจฉัยอีกโรคหนึ่งซึ่งเกิดกับผู้ป่วยอีกกลุ่มหนึ่งที่มีลักษณะคลินิกแตกต่างกัน ตัวอย่างของการใช้ตัวแปรหลายๆ ตัวเพื่อวินิจฉัยแยกโรค 1 ชนิด เช่น การใช้จำนวนนิวโทรฟิลในถุงน้ำดี, การลอยตัวของนิวโทรฟิล, ลักษณะการจับแคลเซียมของนิวโทรฟิล ในการวินิจฉัยแยกนิวโรซิสเป็นโมเลสเตอรอลออกจากนิวโรซิสชนิดอื่นๆ ซึ่งมีวิธีการรักษาแตกต่างกัน หรือการใช้อัตราการหายใจ, อัตราการเดินของชีพจร, อุณหภูมิกาย, forced expiratory volume, การฟังได้ rhonchi, ภาวะ cyanosis ของมือ ฯลฯ ในการวินิจฉัยแยกความรุนแรงของโรคที่คาดว่าจะต้องรับไว้รักษาในโรงพยาบาลหรือไม่ เป็นต้น

การใช้ค่า likelihood ratio จากผลการทดสอบหลายๆ ชนิดเพื่อวินิจฉัยโรคหนึ่ง จึงอาจจะเกิดความคลาดเคลื่อนใน post-test probability ได้ ในกรณีนี้จะต้องใช้วิธีการทางสถิติ เช่น discriminant analysis มาคำนวณค่าน้ำหนักของตัวแปรต่างๆ และตัวแปรที่มีประโยชน์ในการวินิจฉัยแยกโรค ซึ่งเป็นเรื่องที่น่าสนใจและควรปรึกษากับนักสถิติต่อไป

10. ขั้นตอนการประเมินคุณค่าของวิธีการทดสอบที่รายงานในวารสาร

ขั้นตอนต่างๆ ที่ใช้ในการตรวจสอบการประเมินคุณค่าของวิธีการทดสอบเพื่อวินิจฉัยโรค (อาจจะใช้ไม่หมดทุกข้อ) มีดังนี้

9.1 การแปลผลการทดสอบต้องกระทำโดยผู้ที่ไม่ทราบที่มาของตัวอย่าง กล่าวคือ ไม่ทราบลักษณะทางคลินิกของผู้ป่วย หรือไม่ทราบผลการตัดสินว่าผู้ป่วยมีหรือไม่มีโรคตามเกณฑ์มาตรฐาน

ทั้งนี้เพื่อตัดปัญหาเรื่องอคติ (bias) ในการแปลผล ซึ่งในกรณีนี้จัดเป็น

กล่าวคือ ถ้าผู้แปลผลการทดสอบทราบก่อนว่าตัวอย่างที่กำลังจะอ่านผลได้มาจากผู้ป่วยที่มีโรคนั้นแล้ว ก็อาจจะลำเอียงตัดสินผลการทดสอบให้เป็นบวกได้ง่ายขึ้น ส่วนเกณฑ์มาตรฐานและหลักการตัดสินผลการทดสอบต้องมีหลักเกณฑ์ที่ดีและใช้ได้จนเป็นที่ยอมรับทั่วไป นอกจากนี้ผู้ป่วยทุกรายทั้งในกลุ่มเป็นโรคและกลุ่มเปรียบเทียบที่เข้าข่ายในการศึกษาต้องได้รับการทดสอบเหมือนกันทุกราย และวิธีการทดสอบใหม่ต้องมีคุณสมบัติบางอย่างดีเด่นกว่าวิธีที่ใช้เป็นเกณฑ์มาตรฐาน เช่น มีความสะดวก ทำเสร็จรวดเร็ว แม่นยำ มีราคาถูก และทำให้เกิดความเจ็บปวดต่อผู้ถูกทดสอบน้อยกว่า เป็นต้น มิฉะนั้นวิธีใหม่นี้จะไม่เป็นที่นิยมและเลิกใช้ไปเอง นอกจากนี้จะให้ข้อมูลที่สำคัญในด้านอื่นด้วย

9.2 กลุ่มผู้ป่วยที่นำมาศึกษา ควรเป็นกลุ่มที่เป็นปัญหาในการวินิจฉัยโรคทางคลินิก ซึ่งอาจเป็นผู้ที่มีลักษณะทางคลินิกไม่รุนแรง หรือไม่ปฏิบัติตามแบบฉบับ เพราะผู้ป่วยกลุ่มนี้เป็นกลุ่มที่ต้องการข้อมูลเพิ่มเติมจากการทดสอบเพื่อวินิจฉัยแยกโรคให้ชัดเจน อย่างไรก็ตามต้องพิจารณาคุณวิญญูประสงค์ของผู้ทำวิจัยเป็นเกณฑ์ด้วย เช่น ถ้าผู้ทำวิจัยต้องการประเมินความสามารถ (efficacy) ของวิธีทดสอบในสภาพแวดล้อมที่ดีที่สุด ก็อาจจะใช้กลุ่มผู้ป่วยที่มีลักษณะที่เด่นชัดของโรคนั้นมาศึกษาเท่านั้น โดยเปรียบเทียบกับกลุ่มอื่นๆ

9.3 ควรระบุสถานที่ทำการวิจัยในรายงานผลการศึกษา เพราะแสดงถึงความซุกซนของโรคและช่วยในการแปลผลการทดสอบให้กำหนดลักษณะคลินิกของกลุ่มประชากรไว้ก่อนทำการทดสอบเพื่อให้แน่ใจว่า ความแตกต่างของผลการทดสอบในแต่ละกลุ่มเกิดจากโรคเท่านั้น เป็นการตัดอคติในเรื่อง diagnostic suspicion bias

9.4 ถ้ามีการทดสอบซ้ำกับตัวอย่างเดิม ผลที่ได้และการแปลผลควรจะเหมือนเดิม การทดสอบที่ให้ผลเหมือนเดิมจะมีความแน่นอนดีมาก (reproducibility ดี) และจะต้องวัดได้ถูกต้องตามความเป็นจริง (accurate ดี) ถ้าการแปลผลยังได้เหมือนเดิมก็เรียกว่าไม่มี measurement bias ถ้าวิธีทดสอบยังไม่เป็นที่แพร่หลาย ควรแจ้งรายละเอียดของวิธีทดสอบไว้ในรายงานด้วย

9.5 ควรอธิบายถึงที่มาของ "ค่าปกติ" ที่ใช้ในการศึกษา เพราะความหมายของค่าปกติที่ใช้ในการศึกษาแต่ละครั้งโดยผู้วิจัยต่างคนอาจจะไม่เหมือนเดิม และใช้หลักเกณฑ์อย่างไรในการเลือกจุดตัดแบ่งผลทดสอบออกเป็นบวกและลบ

9.6 ถ้าการวินิจฉัยโรคต้องอาศัยการทดสอบหลายๆ วิธีร่วมกัน ควรจะได้ประเมินผลของการทดสอบแต่ละวิธีที่มีต่อการวินิจฉัยขั้นสุดท้ายไว้ด้วย

9.7 การรายงานผลการศึกษาที่ดี ควรบรรยายถึงรายละเอียดของทุกขั้นตอนที่ทำงานวิจัย

9.8 การทดสอบวิธีนั้นได้มีการพิสูจน์ว่ามีประโยชน์ต่อผู้ป่วย ที่ให้ผลบวกจริงและเท็จ ผลบวกรวมหรือเท็จมากน้อยเพียงใด วิธีดังกล่าวนั้นต้องใช้เวลาทำการทดสอบนานหรือสร้างความทุกข์ทรมานแก่ผู้รับการทดสอบมากน้อยเพียงใด

สรุป

บทความนี้ได้กล่าวถึงความรู้พื้นฐานที่จะทำการวิจัยในเรื่องวิธีทดสอบต่างๆ และให้ความรู้ทั่วไปในการแปลผลการทดสอบ การที่มีความรู้พื้นฐานเหล่านี้จะทำให้เข้าใจถึงที่มาของผลบวกเท็จ และผลลบเท็จ และจะได้ระมัดระวังมากขึ้นในการแปลผลการทดสอบ ความรู้พื้นฐานดังกล่าวนี้สามารถนำมาประยุกต์ใช้กับวิธีทดสอบใดๆ ทั้งที่เกี่ยวข้องหรือไม่เกี่ยวข้องกับโรคติดเชื้อ

บรรณานุกรม

1. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 1st ed. Little, Brown and Company: Boston/Toronto 1985.
2. Cebul RD, Beck JR. Biochemical profiles: applications in ambulatory screening and preadmission testing of adults. Ann Intern Med 1987; 106:403-13.
3. Beyond Normality: the predictive value and efficiency of medical diagnoses. In: Galen RS, Gambino SR, 2nd eds. John Wiley and Sons Inc. New York, Chichester, Brisbane, Toronto, Singapore 1975.
4. Clinical Epidemiology: the essentials. In: Fletcher RH, Fletcher SW, Wanger EH, 3rd eds. Williams and Wilkins, Baltimore/London, 1982.

แบบฝึกหัด

หลังจากที่อ่านบทความนี้จบแล้ว ลองทดสอบความสามารถของท่านในการแปลผลการทดสอบวิธี ELISA ในการวินิจฉัยผู้ที่เคยติดเชื้อ human immunodeficiency virus (HIV) ตามโจทย์ที่กำหนดให้ดังนี้

โจทย์

การตรวจหาผู้ที่เคยติดเชื้อโรคเอดส์ (HIV) โดยใช้วิธี ELISA ปรากฏว่า ผู้ป่วยที่เป็นโรคเอดส์ระยะที่ 1 และ 2 ให้ผลบวกร้อยละ 99 ส่วนประชาชนทั่วไปที่ไม่เคยติดเชื้อ HIV ให้ผลลบร้อยละ 99 สมมติว่าความซุกซนของผู้ป่วยโรคเอดส์ระยะที่ 1 และ 2 ในคนไทย พบว่ามีผู้ป่วย 200 รายต่อประชากร 100,000 คน

ถ้ามีหญิงสาวคนหนึ่งขอรับการทดสอบด้วยวิธี ELISA ดังกล่าวนี้และผลการทดสอบให้ผลบวก ท่านจะแปลผลบวกจากการทดสอบอย่างไร?

ถ้าเป็นผู้ที่ให้ผลบวกกับการทดสอบ ELISA มาทำการทดสอบ immunoblot ซึ่งมี sensitivity เท่ากับร้อยละ 99 และ specificity เท่ากับร้อยละ 99 เช่นกัน ท่านจะแปลผลบวกที่ได้จากการทดสอบ immunoblot อย่างไร?

เฉลย

1. ทำการทดสอบ ELISA

	ผู้ป่วยโรคเอดส์ ระยะที่ 1 และ 2	ประชาชนทั่วไป	รวม
ให้ผลบวก	198	998	1,196
ให้ผลลบ	2	98,802	98,804
รวม	200	99,800	100,000

2. ทำการทดสอบ immunoblot

ผู้ที่ให้ผลบวกกับ ELISA

	ผู้ป่วยโรคเอดส์ ระยะที่ 1 และ 2	ประชาชนทั่วไป	รวม
ให้ผลบวก	196	10	206
ให้ผลลบ	2	988	990
รวม	198	998	1,196

โอกาสหญิงสาวรายนี้ที่ให้ผลบวกจะป่วยเป็นโรคเอดส์ระยะ
ที่ 1 หรือ 2 คิดเป็นร้อยละ = $\frac{198}{1,196} \times 100 = 16.6$
นั่นคือ โอกาสที่เธอจะไม่ป่วยเป็นโรคเอดส์คิดเป็นร้อยละ =

สรุป หญิงสาวที่ให้ผลบวกทั้งวิธีทดสอบ ELISA และ
immunoblot รายนี้ มีโอกาสเป็นโรคเอดส์ระยะที่ 1 หรือ 2 คิด
เป็นร้อยละ = $\frac{196}{206} \times 100 = 95$

83.4

“ท่านคำนวณได้เช่นนี้หรือไม่?”