Review article

THAI BULLETIN OF PHARMACEUTICAL SCIENCES
ISSN 2586-8659 (online)

จุดตัดที่เหมาะสมสำหรับการวิเคราะห์เส้นโค้ง Receiver Operating Characteristic (ROC) ในการพัฒนาเครื่องมือนวัตกรรมทางสุขภาพ: กรณีตัวอย่างโดยใช้โปรแกรม STATA

พงษ์เดช สารการ[1,*], ภัทรนันท์ หมั่นพลศรี[1,2]

[1] สาขาวิชาวิทยาการระบาดและชีวสถิติ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น

[2] กลุ่มวิจัยวิทยาการระบาดและป้องกันโรคมะเร็งในภูมิภาคอาเซียน คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น

* ติดต่อผู้นิพนธ์: spongd@kku.ac.th

บทคัดย่อ

การวิเคราะห์เส้นโค้ง receiver operating characteristic (ROC) เป็นวิธีการที่ถูกนำมาใช้ค่อนข้างแพร่หลายในการวินิจฉัยโรค แต่ในการพัฒนานวัตกรรมทางสุขภาพยังคงพบน้อยและไม่ชัดเจนสำหรับแนวปฏิบัติ บทความนี้จึงได้นำเสนอหลักการและตัวบ่งชี้สำหรับการประเมินความถูกต้องของวิธีการวินิจฉัยโรค และการพิจารณาจุดตัดที่เหมาะสม 3 วิธี ได้แก่ วิธี Euclidian's index วิธี Youden's index และวิธี Weighted Number Need to Misdiagnose (Weighted NNM) รวมถึงการประเมินประสิทธิภาพการจำแนกความถูกต้องโดยรวมของเครื่องมือที่เป็นนวัตกรรมทางสุขภาพทั้งแบบวิธีเดียวและแบบเปรียบเทียบสองวิธี ด้วยพื้นที่ใต้โค้ง ROC หรือ area under the curve (AUC) ตัวอย่างถูกสาธิตโดยใช้คำสั่งในโปรแกรม STATA ขณะที่ผลลัพธ์และค่าสถิติที่เกี่ยวข้องยังคงถูกนำเสนอและแปลความหมายด้วยเช่นกัน สุดท้ายเป็นข้อเสนอแนะในการนำวิธีการวิเคราะห์เส้นโค้ง ROC ไปใช้ ซึ่งครอบคลุมทั้งจุดเด่นและข้อจำกัด

คำสำคัญ: กราฟเส้นโค้ง ROC, จุดตัดที่เหมาะสม, ดัชนีของ Youden, พื้นที่ใต้โค้ง AUC, นวัตกรรมทางสุขภาพ

รับต้นฉบับ: 17 พฤศจิกายน 2563; แก้ไข: 23 กุมภาพันธ์ 2564; ตอบรับตีพิมพ์: 25 กุมภาพันธ์ 2564

OPTIMAL CUT-OFF POINTS FOR RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE ANALYSIS IN DEVELOPING TOOLS OF HEALTH INNOVATIONS: EXAMPLE USING STATA

Pongdech Sarakarn[1,*], Pattaranan Munpolsri[1,2]

[1] Epidemiology and Biostatistics Department, Faculty of Public Health, Khon Kaen University, Khon Kaen

[2] ASEAN Caner Epidemiology and Prevention Research Group (ACEP), Faculty of Public Health, Khon Kaen University, Khon Kaen

* Corresponding author: spongd@kku.ac.th

ABSTRACT

Receiver Operating Characteristic (ROC) curve analysis is a method that is used quite widely in Diagnosis of disease, but for use in the development of health innovations it is rarely seen, and its possible application in this area remains unclear. This article presents the principles and indicators that are used for assessing the accuracy of diagnostic tests, and considers three optimal cut-off point methods, including - Euclidian's index, Youden's index and Weighted Number Needed to Misdiagnose (Weighted NNM). In addition, this work assesses the overall discriminant accuracy of the tests, for their potential application in health innovation - assessing each method individually and comparatively, using the area below the ROC curve (AUC). Examples were demonstrated using commands in STATA software programme. The results and related statistics were presented and interpreted as well. Finally, the implications of using ROC analysis were identified, which cover both the advantages and limitations of the ROC analysis.

Keywords: receiver operating characteristics (ROC), optimal cut-off point, Youden index, area under the curve (AUC), health innovation

Received: 17 November 2020; Revised: 23 February 2021; Accepted: 25 February 2021

บทนำ

นวัตกรรมทางสุขภาพ (health innovations) ของประเทศในปัจจุบัน ได้ถูกส่งเสริมให้พัฒนาและมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง โดยเฉพาะนวัตกรรมทางสุขภาพในรูปแบบมาตรวัด หรือ แบบประเมิน ซึ่งพบมากในงานวิจัยทางวิทยาศาสตร์สุขภาพบางสาขา เช่น สาขาทางจิตเวช สาขาทางการพยาบาลและสาธารณสุข เป็นต้น ภายใต้เงื่อนไขและข้อจำกัดของวิธีการมาตรฐาน (gold standard) เดิมที่มีอยู่ในทางปฏิบัติ เช่น เครื่องมือมีจำนวนน้อยไม่ครอบคลุม มีค่าใช้จ่ายสูง กระบวนการอาจมีความซับซ้อนและใช้เวลานาน หรือ จำเป็นต้องใช้ผู้เชี่ยวชาญเฉพาะด้าน ซึ่งอาจมีไม่เพียงพอ เป็นต้น จากข้อจำกัดดังกล่าว การสร้างนวัตกรรมทางสุขภาพ โดยเฉพาะเครื่องมือที่ใช้สำหรับการประเมินเพื่อตรวจคัดกรอง จึงถูกพัฒนาขึ้นและเปรียบเสมือนตัวแทน (surrogate) ของวิธีการมาตรฐานเดิมที่มีอยู่[1] เพื่อนำมาใช้บ่งชี้ หรือ จำแนกโอกาสการเกิดโรค รวมถึงความผิดปกติบางอย่าง หรือ จำแนกโอกาสเสี่ยงที่เกิดขึ้นจากกลุ่มตัวอย่างที่ศึกษาว่า “มีความถูกต้องมากน้อยเพียงใด?” เมื่อเทียบเคียงกับวิธีการมาตรฐานเดิมที่มีอยู่ โดยในทางปฏิบัติ ผลลัพธ์ที่คาดหวังและถูกนำมาใช้ในการแปลความหมาย เพื่อประเมินความถูกต้อง (accuracy) ของเครื่องมือ ส่วนใหญ่มุ่งเน้นและให้ความสำคัญกับผลลัพธ์ในรูปแบบแจงนับ 2 ค่า (dichotomous outcome)[2] เช่น “เกิดโรค/ไม่เกิดโรค” “ผิดปกติ/ปกติ” หรือ “เสี่ยง/ไม่เสี่ยง” เป็นต้น ดังนั้น เครื่องมือที่ให้ค่าผลลัพธ์แบบอื่น จึงจำเป็นต้องแปลงผลลัพธ์ให้อยู่ในรูปแบบแจงนับ ด้วยวิธีการวิเคราะห์เส้นโค้ง Receiver Operating Characteristics (ROC) โดยเฉพาะเครื่องมือที่ให้ค่าผลลัพธ์แบบต่อเนื่อง (continuous outcome) เช่น ค่าคะแนนที่ได้จากแบบวัดต่าง ๆ เป็นต้น ซึ่งแต่ละค่าของผลลัพธ์แบบต่อเนื่องที่เป็นไปได้ดังกล่าว จะถูกนำมากำหนดเป็นจุดตัด (cut-off point) เพื่อจำแนกผลลัพธ์ออกเป็น 2 ค่าแบบแจงนับ ได้แก่ 0=“ต่ำกว่าค่าจุดตัด” และ 1=“มากกว่า หรือเท่ากับจุดตัด” และนำผลลัพธ์ที่ได้มาเทียบกับผลลัพธ์จากวิธีการมาตรฐาน เพื่อประเมินค่าความไว (sensitivity) และความจำเพาะ (specificity) ในแต่ละจุดตัดของผลลัพธ์แบบต่อเนื่องที่เป็นไปได้ทั้งหมด พร้อมแสดงเป็นกราฟเส้นโค้ง ROC จากนั้นจึงพิจารณาและตัดสินใจเลือกค่าจุดตัดที่ให้ค่าความไวและความจำเพาะที่เหมาะสม (optimal cut-off point) เพื่อนำไปใช้ประเมิน หรือ เปรียบเทียบความถูกต้องกับเครื่องมืออื่นและนำไปสู่การตัดสินใจเลือกเครื่องมือที่มีประสิทธิภาพสูงสุดไปใช้ประโยชน์ในทางปฏิบัติต่อไป อย่างไรก็ตาม การพิจารณาเพื่อกำหนดจุดตัดที่เหมาะสมดังกล่าว ยังสามารถกำหนดจุดตัดได้มากกว่าหนึ่งค่า ขึ้นกับวัตถุประสงค์และการนำไปใช้ประโยชน์ในทางปฏิบัติ เช่น การกำหนดจุดตัด 2 ค่า เพื่อจำแนกคนออกเป็น 3 กลุ่ม ได้แก่ (1) กลุ่มที่มีความเสี่ยงต่ำ เพื่อประโยชน์ในการแนะนำและเฝ้าระวัง (2) กลุ่มที่มีความเสี่ยงปานกลาง เพื่อประโยชน์ในการส่งต่อสำหรับการตรวจหรือ การวินิจฉัยเพิ่มเติมด้วยวิธีการที่ให้ผลแม่นยำมากขึ้น และ (3) กลุ่มที่มีความเสี่ยงสูง เพื่อประโยชน์ในการส่งต่อสำหรับการดูแลรักษาให้มีประสิทธิภาพตามกระบวนการที่มีอยู่ เป็นต้น หรือ การกำหนดจุดตัดแบบหนึ่งค่า ภายใต้เงื่อนไขที่แตกต่างกันมากกว่าหนึ่งทางเลือก เช่น การกำหนดจุดตัดที่เหมาะสมแบบหนึ่งค่า ภายใต้ทางเลือกอัตราความชุกของโรคที่แตกต่างกัน เป็นต้น

ซึ่งจากแนวทางดังกล่าว ได้มีการนำเสนอวิธีการพิจารณาค่าจุดตัดที่เหมาะสมและนำมาใช้ประยุกต์กับการพัฒนาเครื่องมือต่าง ๆ ค่อนข้างมากในต่างประเทศ โดยเฉพาะสาขาทางวิทยาศาสตร์สุขภาพ ยกตัวอย่างเช่น การพัฒนาจุดตัดที่เหมาะสมของค่าดัชนีมวลกาย (BMI)[3], แบบวัดคุณภาพชีวิตขององค์การอนามัยโลกชุดย่อ (WHOQOL-bref)[4], แบบประเมินภาวะโภชนาการด้วย Minimal Eating Observation and Nutrition Form-Version II (MEONF-II)[5], แบบประเมินความเสี่ยงต่อการพลัดตกหกล้ม (Morse Fall Scale)[6], ตัวบ่งชี้ cardiometabolic risk[7] และแบบประเมิน sensitive skin[8]

เป็นต้น ขณะที่ในประเทศไทย พบการนำวิธีการพิจารณาค่าจุดตัดที่เหมาะสมด้วยการวิเคราะห์เส้นโค้ง ROC มาใช้ส่วนใหญ่ในงานวิจัยทางคลินิค หรือ การวินิจฉัยโรค แต่การนำมาประยุกต์ใช้ในการพัฒนางานวิจัยนวัตกรรมทางสุขภาพ โดยสังเกตจากบทความวิชาการ หรือ บทความวิจัยที่เผยแพร่ พบว่า ยังจำกัดและมีจำนวนน้อย โดยเฉพาะบทความวิชาการที่นำเสนอวิธีการกำหนดค่าจุดตัดที่เหมาะสมของการวิเคราะห์เส้นโค้ง ROC กรณีเครื่องมือที่มีผลลัพธ์แบบต่อเนื่อง โดยใช้โปรแกรมสำเร็จรูปทางสถิติที่มีอยู่

ดังนั้นบทความวิชาการนี้ จึงได้นำวิธีการวิเคราะห์เส้นโค้ง ROC ซึ่งเป็นวิธีประเมินความถูกต้องสำหรับวิธีการวินิจฉัยโรค (diagnostic test) กรณีผลลัพธ์แบบต่อเนื่องมาประยุกต์ใช้ในการประเมินความถูกต้องของเครื่องมือ ซึ่งเป็นนวัตกรรมทางสุขภาพ โดยมีวัตถุประสงค์เพื่อนำเสนอวิธีการประเมินค่าจุดตัดที่เหมาะสมของเครื่องมือกรณีผลลัพธ์แบบต่อเนื่อง 3 วิธี ได้แก่ วิธี Euclidian's index วิธี Youden's index และวิธี Weighted Number Need to Misdiagnose (Weighted NNM) และวิธีการเปรียบเทียบเครื่องมือ จากพื้นที่ใต้โค้งของเส้นกราฟ ROC (area under the curve; AUC) ทั้งกรณีวิธีเดียวและเปรียบเทียบสองวิธี โดยใช้คำสั่งในโปรแกรม STATA เวอร์ชั่น 15 ประกอบด้วยหัวข้อหลักการและตัวบ่งชี้สำหรับการประเมินความถูกต้องของวิธีการวินิฉัยโรค การประเมินความถูกต้องของวิธีการวินิจฉัยโรคภายใต้การพิจารณาจุดตัดที่เหมาะสม การเปรียบเทียบความถูกต้องของวิธีการวินิจฉัยโรคด้วยการพิจารณาพื้นที่ใต้โค้งของกราฟ ROC (AUC) กรณีตัวอย่างการวิเคราะห์เส้นโค้ง ROC ด้วยโปรแกรม STATA รวมถึงบทสรุปและข้อเสนอแนะเพื่อการนำไปใช้

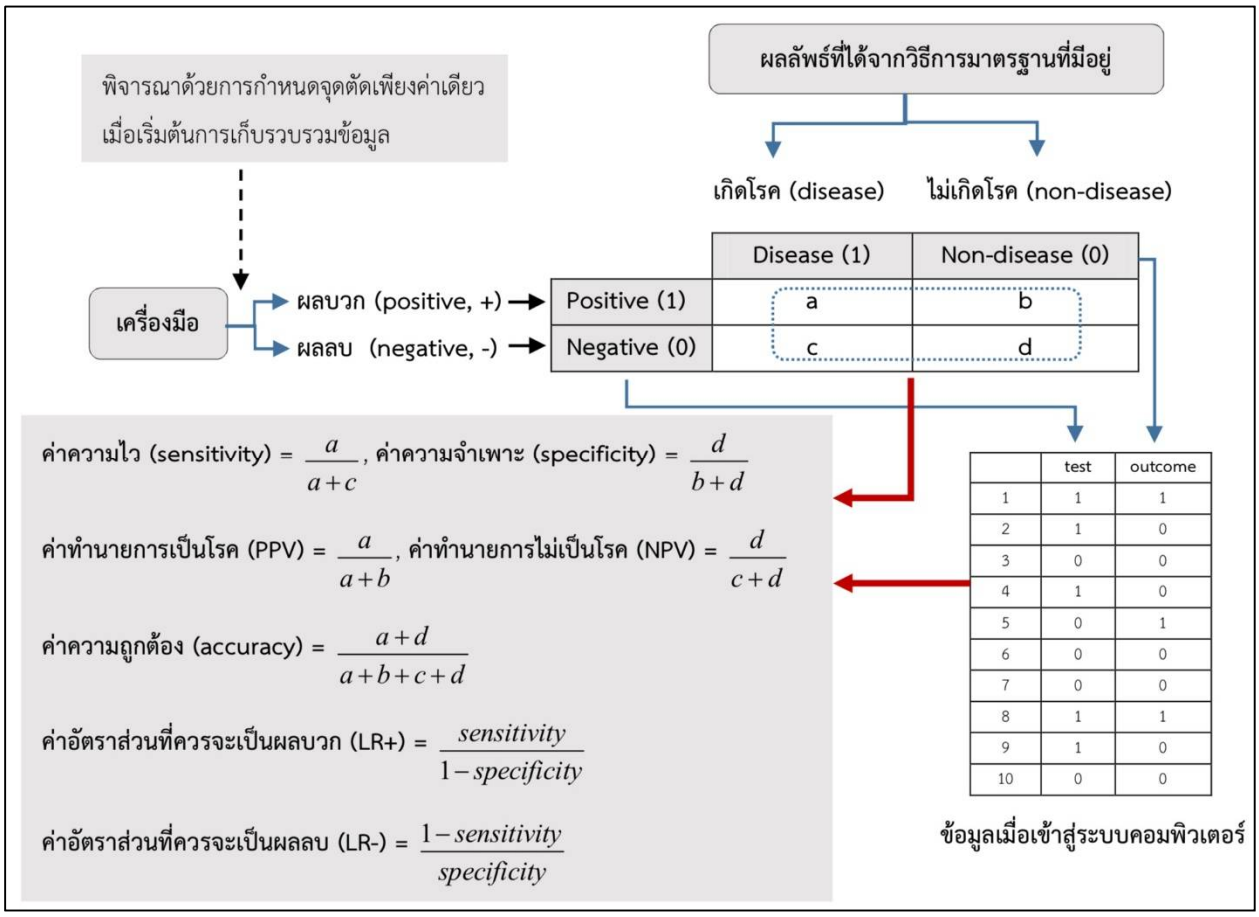
หลักการและตัวบ่งชี้สำหรับการประเมินความถูกต้องของวิธีการวินิจฉัยโรค

วิธีการวินิจฉัยโรค ถือเป็นเครื่องมือหนึ่งที่มีความสำคัญและมีบทบาทอย่างมากต่อระบบการให้บริการทางการแพทย์และสาธารณสุขของประเทศ โดยผลลัพธ์ที่ได้จากการประเมินด้วยวิธีการวินิจฉัยโรค ไม่เพียงแต่สามารถนำมาใช้เพื่อยืนยันการเกิดโรคในกลุ่มผู้ป่วยได้เท่านั้น แต่ยังสามารถนำมาใช้เพื่อจำแนกคนที่ไม่เกิดโรคออกจากกลุ่มผู้ป่วยที่เกิดโรคได้อีกด้วย ซึ่งในทางปฏิบัติผลลัพธ์ที่เป็นไปได้จากวิธีการวินิจฉัยโรค สามารถให้ค่าผลลัพธ์เป็นไปได้ ทั้งแบบแจงนับ 2 ค่า ได้แก่ ค่าผลบวก (positive, +) และค่าผลลบ (negative, -) และแบบต่อเนื่อง เช่น ระดับสารบิลิรูบิน (bilirubin level) เป็นต้น ในการประเมินความถูกต้องของวิธีการวินิจฉัยโรค การเลือกวิธีการที่เหมาะสมขึ้นกับลักษณะข้อมูลผลลัพธ์ของแต่ละวิธีที่นำมาใช้ โดยในทางปฏิบัติ สำหรับวิธีการที่มีลักษณะข้อมูลผลลัพธ์แบบแจงนับ 2 ค่า สามารถพิจารณาความถูกต้องได้ ด้วยการนำผลลัพธ์เทียบกับผลลัพธ์จากวิธีการมาตรฐานเดิมและแสดงเป็นความถี่ในตาราง 2x2 (ในmachine learning; ML ของระบบปัญญาประดิษฐ์ หรือ artificial intelligent; AI เรียกว่า "confusion matrix"[9]) ภายใต้จุดตัดที่พิจารณาจากเครื่องมือเพียงค่าเดียวและนำมาคำนวณเป็นตัวบ่งชี้ที่เกี่ยวข้อง ได้แก่ ค่าความไว ค่าความจำเพาะ ค่าทำนายการเป็นโรค (positive predictive value, PPV) ค่าทำนายการไม่เป็นโรค (negative predictive value, NPV) ค่าความถูกต้อง (accuracy) ค่าอัตราส่วนที่ควรจะเป็นของผลบวก (likelihood ratio positive, LR+) และค่าอัตราส่วนที่ควรจะเป็นของผลลบ (likelihood ratio negative, LR-) (ดังรูปที่ 1)

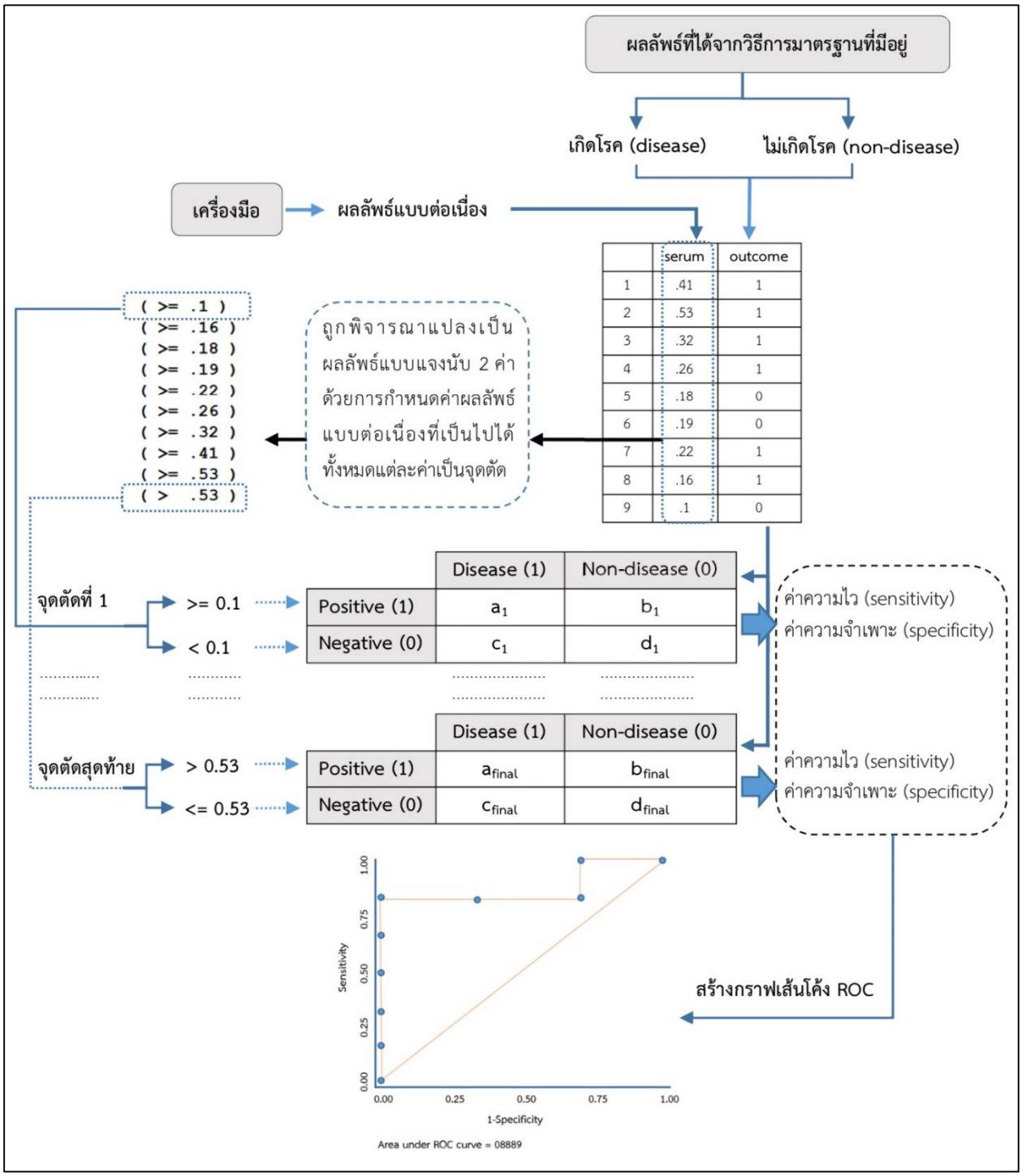
สำหรับวิธีการวินิจฉัยโรคที่มีผลลัพธ์แบบต่อเนื่องยังคงสามารถพิจารณาความถูกต้องได้เช่นเดียวกับกรณีผลลัพธ์แบบแจงนับ 2 ค่า ดังรูปที่ 1 ด้วยการนำผลลัพธ์ที่ได้เทียบกับผลลัพธ์จากวิธีการมาตรฐานเดิมและแสดงเป็นความถี่ในตาราง 2x2 แต่เนื่องจากผลลัพธ์ที่ได้ ให้ค่าแบบต่อเนื่องและมีความเป็นไปได้ ตั้งแต่สองค่าขึ้นไป จึงทำให้การพิจารณากำหนดค่าจุดตัด เพื่อแปลงผลลัพธ์ที่ได้จากแบบต่อเนื่องเป็นแบบแจงนับ 2 ค่า ต้องอาศัยการ

จัดเรียงและจำแนก ขึ้นกับแต่ละค่าของผลลัพธ์แบบต่อเนื่องที่เป็นไปได้ทั้งหมด ดังนั้นในการคำนวณค่าความไวและความจำเพาะ หรือ ค่าบ่งชี้อื่นที่เกี่ยวข้อง จึงถูกพิจารณาและผันแปรไปตามจุดตัดทุกค่าที่เป็นไปได้ทั้งหมด หลังจากนั้นจึงนำค่าความไวและความจำเพาะของแต่ละจุดตัดที่เป็นไปได้ดังกล่าว มาสร้างเป็นกราฟเส้นโค้ง โดยกำหนดให้แกน X แทน “1-specificity” หรือ “false positive fraction (FPF)” และแกน Y แทน ค่าความไว (sensitivity) หรือ “true positive fraction (TPF)” และเรียกกราฟเส้นโค้งนี้ว่า “Receiver Operating Characteristic (ROC) curve”[10] (ดังรูปที่ 2)

จากรูปที่ 2 พบว่า เมื่อมีการนำวิธีการวินิจฉัยโรคที่ให้ค่าผลลัพธ์แบบต่อเนื่องมาประเมินความถูกต้อง จึงจำเป็นต้องอาศัยวิธีการวิเคราะห์เส้นโค้ง ROC ด้วยการกำหนดให้ผลลัพธ์แบบต่อเนื่องแต่ละค่าที่เป็นไปได้ทั้งหมดเป็นจุดตัด เพื่อแบ่งข้อมูลออกเป็นสองส่วนแบบแจงนับ นั่นคือ เมื่อผลลัพธ์แบบต่อเนื่องมีค่าต่ำกว่า (<) จุดตัดที่กำหนด จะถูกจัดให้มีผลลัพธ์เป็น “ผลลบ (negative, -)” แทนด้วยสัญลักษณ์เลขศูนย์ (0) ในระบบคอมพิวเตอร์และเมื่อผลลัพธ์แบบต่อเนื่องมีค่ามากกว่าหรือ เท่ากับ (≥) จุดตัดที่กำหนด จะถูกจัดให้มีผลลัพธ์เป็น “ผลบวก (positive, +)” แทนด้วยสัญลักษณ์ตัวเลขหนึ่ง (1) ในระบบคอมพิวเตอร์ จากนั้นจึงนำผลลัพธ์แบบแจงนับที่ได้เทียบกับผลลัพธ์จากวิธีการมาตรฐานเดิมและแจกแจงออกมาเป็นความถี่ในตาราง 2x2 จำแนกตามแต่ละจุดตัดที่เป็นไปได้ทั้งหมดของผลลัพธ์แบบต่อเนื่อง ดังนั้นจึงส่งผลให้ค่าความไวและความจำเพาะที่ได้ มีความแตกต่างกันขึ้นกับจุดตัดของแต่ละค่าผลลัพธ์ที่เกิดขึ้น



รูปที่ 1 การพิจารณาตัวบ่งชี้ในการประเมินความถูกต้องของวิธีการวินิจฉัยโรคที่มีผลลัพธ์แบบแจงนับ 2 ค่า



รูปที่ 2 การพิจารณาตัวบ่งชี้ในการประเมินความถูกต้องของวิธีการวินิจฉัยโรคที่มีผลลัพธ์แบบต่อเนื่อง

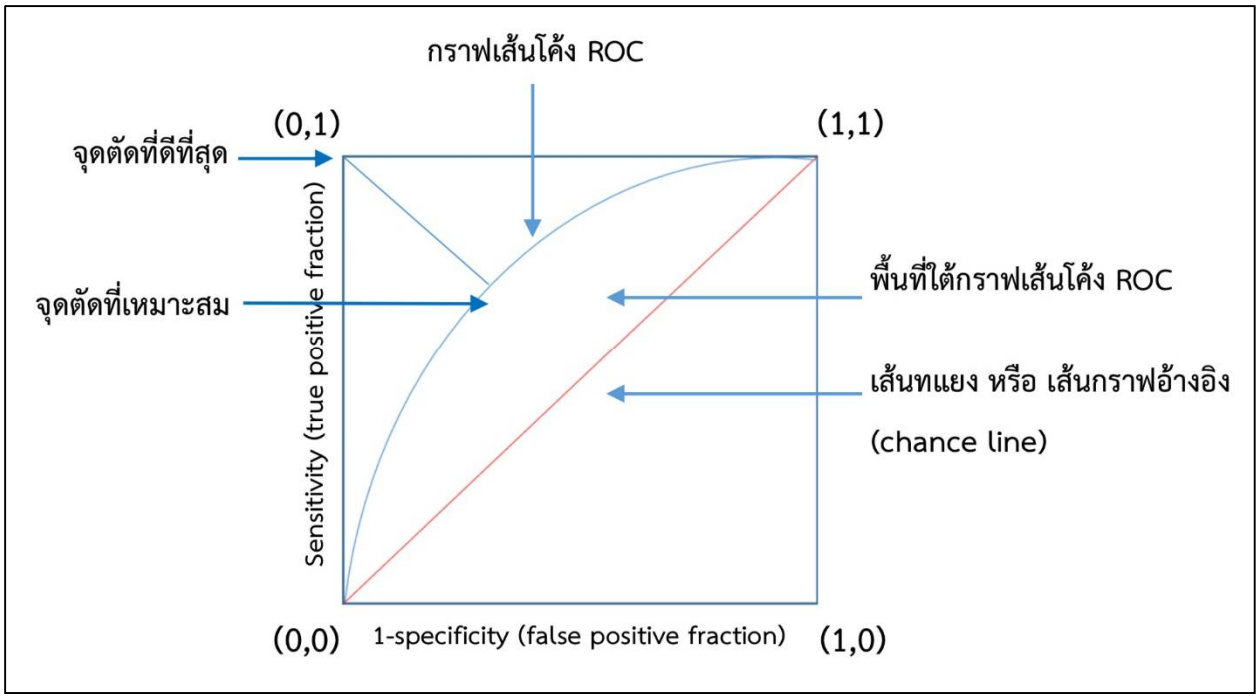
และเมื่อนำมาสร้างเป็นกราฟเส้นโค้ง ROC จึงสามารถสะท้อนได้อย่างชัดเจนถึงการเปลี่ยนแปลงที่เกิดขึ้นของค่าความไวและความจำเพาะ เมื่อค่าจุดตัดที่ถูกพิจารณาผันแปรไป[11] ซึ่งจากผลดังกล่าว จึงทำให้นักวิจัยสามารถนำผลลัพธ์ที่ได้จากกราฟเส้นโค้ง ROC มาใช้ประโยชน์สำหรับการประเมินวิธีการวินิจฉัยโรค โดยเฉพาะใน 2 ประเด็นหลัก ได้แก่ (1) การประเมินความถูกต้องของวิธีการวินิจฉัยโรคภายใต้การพิจารณาจุดตัดที่เหมาะสม และ (2)

การเปรียบเทียบความถูกต้องของวิธีการวินิจฉัยโรคด้วยการพิจารณาพื้นที่ใต้โค้งของกราฟ ROC (AUC) ทั้งในกรณีวิธีเดียวและกรณีเปรียบเทียบสองวิธี

การประเมินความถูกต้องของวิธีการวินิจฉัยโรคภายใต้การพิจารณาจุดตัดที่เหมาะสม

การประเมินความถูกต้องของวิธีการวินิจฉัยโรคภายใต้การพิจารณาจุดตัดที่เหมาะสม ถือเป็นอีกหนึ่งประโยชน์ที่สำคัญและสามารถพิจารณาได้จากการวิเคราะห์ด้วยกราฟเส้นโค้ง ROC กรณีผลลัพธ์แบบต่อเนื่อง เพราะกราฟเส้นโค้ง ROC จะแสดงความสัมพันธ์ระหว่างแกน X ซึ่งแทน “1-specificity” หรือ ส่วนของผลบวกปลอม “false positive fraction (FPF)” กับ แกน Y ซึ่งแทน“sensitivity” หรือ ส่วนของผลบวกจริง “true positive fraction (TPF)” โดยจุดบนกราฟเส้นโค้ง ROC ที่เกิดขึ้น หมายถึง จุดตัด (cut-off point หรือ threshold) ที่เป็นไปได้ระหว่างความสัมพันธ์ของทั้งสองแกนดังกล่าว ดังนั้นจุดตัดที่ดีที่สุด (perfectly cut-off point) จึงหมายถึง จุดตัดที่ให้ค่าความไว หรือ ค่าผลบวกจริง เท่ากับ 1 และให้ค่าผลบวกปลอม เท่ากับ 0 หรือ จุดตัด ณ ตำแหน่ง (0,1) นั่นเอง แต่ในความเป็นจริง วิธีการวินิจฉัยโรคที่ถูกนำมาใช้และมีจุดตัดที่ดีที่สุดดังกล่าว มีความเป็นไปได้น้อยมาก ดังนั้นจุดตัดที่เหมาะสม หรือ optimal cut-off point จึงถูกนำมาพิจารณาและเป็นทางเลือก ซึ่งหมายถึง จุดตัดที่สามารถให้ค่าความไว หรือ ค่าผลบวกจริงได้มากที่สุดและขณะเดียวกันก็ให้ค่าผลบวกปลอมน้อยที่สุดเช่นเดียวกัน ในทางปฏิบัติ จุดตัดที่เหมาะสมดังกล่าว จึงสามารถพิจารณาได้จากจุดตัดบนกราฟเส้นโค้ง ROC ที่เข้าใกล้มุมบนด้านซ้ายมือ หรือ เข้าใกล้จุดในตำแหน่ง (0,1) มากที่สุด[12] (ดังรูปที่ 3)

ดังนั้นการพิจารณาจุดตัดที่เหมาะสมของวิธีการวินิจฉัยโรค จึงถือเป็นประเด็นสำคัญที่นักวิจัยจำเป็นต้องคำนึงถึง เพราะมีผลโดยตรงต่อการตัดสินใจเลือกวิธีการวินิจฉัยโรคที่ได้ให้มีความถูกต้องและมีความแม่นยำในการจำแนกคนออกเป็นกลุ่มมากที่สุด ซึ่งในทางปฏิบัติ วิธีการพิจารณาจุดตัดที่เหมาะสมมีได้หลายวิธีขึ้นกับมุมมองของนักวิจัยที่แตกต่างกัน เช่น วิธีการพิจารณาจุดตัดที่เหมาะสม



รูปที่ 3 ลักษณะทั่วไปของกราฟเส้นโค้ง ROC

ภายใต้มุมมองความน่าจะเป็นและประเด็นทางคลินิก[2] หรือวิธีการพิจารณาภายใต้มุมมองความผันแปรของการแจกแจงปกติแบบ binormal[12] รวมถึงวิธีพิจารณาจุดตัดที่เหมาะสม ภายใต้ข้อมูลสนับสนุนอื่นที่สำคัญเพิ่มเติม เช่น อัตราความชุกของโรค หรือ ค่าใช้จ่ายที่เกิดขึ้น เป็นต้น แต่อย่างไรก็ตาม สำหรับวิธีพิจารณาจุดตัดที่เหมาะสมเพื่อประเมินความถูกต้องของเครื่องมือ ซึ่งเป็นนวัตกรรมทางสุขภาพในครั้งนี้ เพื่อให้เกิดความครอบคลุมในการนำไปใช้ปฏิบัติจริง ทั้งการกำหนดจุดตัดแบบค่าเดียวและแบบการนำข้อมูลสนับสนุนอื่นที่เกี่ยวข้องมาพิจารณาประกอบ ในบทความนี้ จึงได้นำเสนอวิธีการพิจารณาจุดตัดที่เหมาะสมของเครื่องมือที่เป็นนวัตกรรมสุขภาพกรณีผลลัพธ์แบบต่อเนื่อง ประกอบด้วย 3 วิธีการ[13] ได้แก่ วิธี Euclidean index วิธี Youden's index และวิธี Weighted Number Need to Misdiagnose (Weighted NNM) โดยมีรายละเอียดดังนี้

1. การพิจารณาจุดตัดที่เหมาะสมด้วยวิธี Euclidean's index บางครั้งอาจเรียกว่า วิธีเข้าใกล้จุด (0,1) หรือ "closest to (0,1)" แทนด้วยสัญลักษณ์ " i " เป็นวิธีการกำหนดจุดตัดแบบค่าเดียว โดยถือเป็นวิธีหาระยะทางเรขาคณิต (geometric distance) นั่นคือ เป็นการพิจารณาระยะห่างระหว่างจุด (0,1) ซึ่งถือเป็นจุดที่มีค่าความไวเท่ากับ 1 และมีค่าผลบวกปลอมเท่ากับ 0 กับจุด $(1-sensitivity, 1-specificity)$ ที่เป็นไปได้บนเส้นกราฟ ROC ภายใต้แนวคิดการเทียบเคียงวิธีหาระยะทางเรขาคณิตของรูปสามเหลี่ยมมุมฉากและมีสูตรในการพิจารณาดังนี้

$$c^2 = (1-sensitivity)^2 + (1-specificity)^2 \quad \text{หรือ}$$
$$c = \sqrt{(1-sensitivity)^2 + (1-specificity)^2}$$

เมื่อ c = ระยะห่างระหว่างจุด (0,1) กับจุด $(sensitivity, 1-specificity)$ ที่เป็นไปได้บนเส้นกราฟ ROC (ดังรูปที่ 4)

เกณฑ์การพิจารณาจุดตัดที่เหมาะสมของวิธี Euclidean index ได้แก่

> จุดตัดที่มีค่าระยะห่างระหว่างจุด (0,1) กับจุด $(sensitivity, 1-specificity)$ ที่เป็นไปได้บนเส้น กราฟ ROC เท่ากับ ***ค่าน้อยที่สุด***

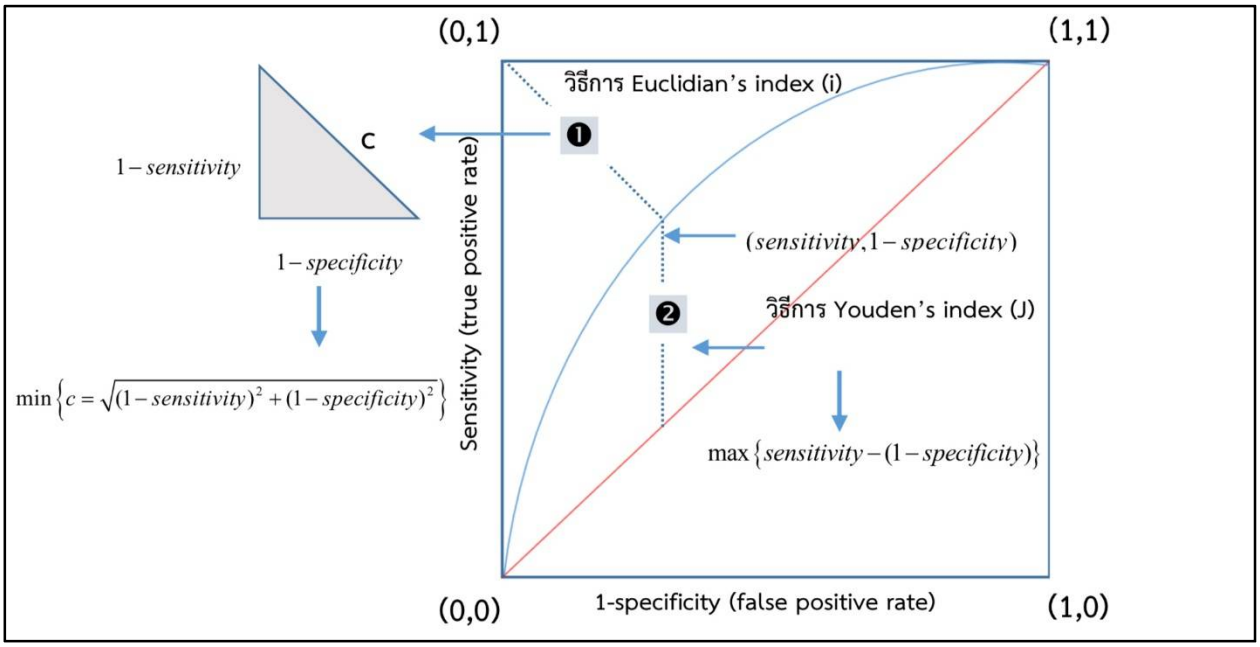
2. การพิจารณาจุดตัดที่เหมาะสมด้วยวิธี Youden's index แทนด้วยสัญลักษณ์ " J " เป็นวิธีการกำหนดจุดตัดแบบค่าเดียว โดยเป็นการพิจารณาระยะห่างระหว่างจุด $(sensitivity, 1-specificity)$ ที่เป็นไปได้บนเส้นกราฟ ROC กับเส้นทแยง หรือ chance line ในแนวดิ่ง (ดังรูปที่ 4) ซึ่งมีสูตรในการพิจารณาดังนี้

$$J = sensitivity + specificity - 1 \quad \text{หรือ}$$
$$J = sensitivity - (1-specificity)$$

เกณฑ์การพิจารณาจุดตัดที่เหมาะสมของวิธี Youden's index ได้แก่

> จุดตัดที่มีค่าระยะห่างระหว่างจุด $(sensitivity, 1-specificity)$ ที่เป็นไปได้บนเส้นกราฟ ROC กับเส้นทแยง หรือ chance level ในแนวดิ่ง เท่ากับ ***ค่ามากที่สุด***

อย่างไรก็ตาม แม้การพิจารณาจุดตัดที่เหมาะสมด้วยวิธี Euclidean index และวิธี Youden's index ข้างต้น จะเป็นวิธีการที่ง่ายและตรงไปตรงมา อีกทั้งยังถูกยอมรับและนำมาใช้ค่อนข้างแพร่หลาย แต่ในทางปฏิบัติทั้งสองวิธีดังกล่าว ถูกพิจารณาภายใต้แนวคิดของตำแหน่งเชิงคณิตศาสตร์เพียงด้านเดียว ซึ่งอาจไม่สอดคล้องกับบริบทของการนำเครื่องมือไปใช้จริง ดังนั้นจึงได้มีการพัฒนาวิธีการเพิ่มเติม ด้วยการนำข้อมูลสนับสนุนอื่นที่มีล่วงหน้าและเกี่ยวข้องมาพิจารณาประกอบ ภายใต้แนวคิดการวิเคราะห์แบบเบย์เซียน (Bayesian analysis) เพื่อให้การกำหนดจุดตัดที่ได้มีความเหมาะสมเพิ่มมากขึ้น โดยเฉพาะ



รูปที่ 4 วิธีการพิจารณาจุดตัดที่เหมาะสมด้วยวิธี Euclidian's index และวิธี Youden's index

ในประเด็นความสอดคล้องกับบริบทของการเกิดโรคและความคุ้มค่าในการดำเนินการ ดังรายละเอียดในวิธีการถัดไป

3. การพิจารณาจุดตัดที่เหมาะสมด้วยวิธี Weight Number Need to Misdiagnose แทนด้วยสัญลักษณ์ " Weighted NNM " เป็นวิธีการที่พัฒนามาจากแนวคิดการวิเคราะห์แบบ Bayesian ซึ่งให้ความสำคัญกับการนำข้อมูลที่มีอยู่ก่อนล่วงหน้ามาใช้ประโยชน์ในการคำนวณหาโอกาสความน่าจะเป็นของการเกิดโรคหลังการทดสอบด้วยเครื่องมือ (Post-test odds of D+) โดยวิธีการนี้ได้คำนึงถึงข้อมูลล่วงหน้าที่นำมาพิจารณาประกอบด้วย **(1) ข้อมูลโอกาสความน่าจะเป็นของการเกิดโรคก่อนการทดสอบด้วยเครื่องมือ** (Pre-test odds of D+) แทนด้วยสัญลักษณ์ " pr " ซึ่งในทางปฏิบัติสามารถพิจารณาข้อมูลความชุกของการเกิดโรค แทนได้[14] และ **(2) ข้อมูลความคุ้มค่า** ซึ่งพิจารณาจากค่าใช้จ่าย (cost) ของผลกระทบที่เกิดขึ้นจากค่าผลลบปลอม (False negative, FN) และค่าผลบวกปลอม (False positive, FP) ผ่านตัวชี้วัดที่เรียกว่า "Number Needed to Misdiagnoses, NNM"[15] ซึ่งหมายถึง จำนวนผู้ป่วยที่ถูกตรวจด้วยเครื่องมือ แล้วมีหนึ่งรายในจำนวนดังกล่าวที่ถูกวินิจฉัยผิด เช่น หาก NNM=15 หมายความว่า มีผู้ป่วย 1 ใน 15 รายที่ถูกตรวจด้วยเครื่องมือทดสอบแล้ว ถูกวินิจฉัยผิด (ซึ่งอาจเป็นไปได้ทั้งแบบ FN หรือ แบบ FP) นั่นแสดงว่า เครื่องมือที่ให้ค่า NNM สูง ย่อมถือเป็นเครื่องมือที่มีความถูกต้องและใกล้เคียงกับวิธีการมาตรฐาน แต่อย่างไรก็ตามในทางปฏิบัติ ค่าใช้จ่ายที่เกิดขึ้นจากผลกระทบของ FN และ FP มีความแตกต่างกันไปตามบริบทของโรคและการนำไปใช้ ดังนั้นการถ่วงน้ำหนักด้วยจำนวน C เท่าของค่าใช้จ่ายระหว่างผลกระทบ FN และ FP จึงถูกพิจารณาและนำมากำหนดเป็นสูตรในการคำนวณเพื่อพิจารณาค่าจุดตัดที่เหมาะสมได้ดังนี้[1]

$$\text{weighted NNM} = \frac{1}{(C \times FN) + FP} = \frac{1}{(C \times pr(1-sen)) + (1-pr)(1-spec)}$$

เมื่อ C = จำนวนเท่าของค่าใช้จ่ายจากผลกระทบ FN เมื่อเทียบกับ FP เช่น C=3 แสดงว่าค่าใช้จ่ายจากผลกระทบ FN มีมากเป็น 3 เท่าเมื่อเทียบกับ FP

pr = โอกาสความน่าจะเป็นของการเกิดโรคก่อนการทดสอบด้วยเครื่องมือ หรือ ความชุกของการเกิดโรค เช่น ในการศึกษาเกี่ยวกับความดันโลหิตสูง พบว่า ผู้ป่วยเบาหวานจำนวน 200 ราย ถูกวินิจฉัยว่า เป็นโรคความดันโลหิตสูง จำนวน 78 ราย จะได้ความชุกของการเกิดโรคเป็น 78/200 = 0.39 เป็นต้น

sen = ค่าความไวที่ได้จากจุดตัดที่เป็นไปได้ทั้งหมดบนค่าผลลัพธ์แบบต่อเนื่อง

spec = ค่าความจำเพาะที่ได้จากจุดตัดที่เป็นไปได้ทั้งหมดบนค่าผลลัพธ์แบบต่อเนื่อง

เกณฑ์การพิจารณาจุดตัดที่เหมาะสมของวิธี Weighted NNM ได้แก่

จุดตัดที่ให้ค่า Weighted NNM เท่ากับ ***ค่ามากที่สุด***

การเปรียบเทียบความถูกต้องของวิธีการวินิจฉัยโรคด้วยการพิจารณาพื้นที่ใต้โค้งของกราฟ ROC

พื้นที่ใต้โค้ง ROC หรือ area under the curve (AUC) บางครั้งเรียก “C-statistics” ถือเป็นการประเมินประสิทธิภาพการจำแนกความถูกต้องโดยรวมของเครื่องมือ ซึ่งเครื่องมือที่มีประสิทธิภาพการจำแนกความถูกต้องได้สมบูรณ์ (perfect discrimination) จะมีค่าพื้นที่ AUC เท่ากับ 1 และจะลดหลั่นลงมาจนถึงต่ำกว่า 0.5 ซึ่งถือว่า ไม่มีประสิทธิภาพในการจำแนกความถูกต้องได้ (no discrimination) โดยในทางปฏิบัติ การวิเคราะห์พื้นที่ใต้โค้งของกราฟ ROC หรือ AUC ถูกพิจารณาแบ่งออกเป็น 2 แนวทาง ประกอบด้วย **(1) วิธีการวิเคราะห์แบบนอนพาราเมตริก** เป็นวิธีพิจารณาพื้นที่ AUC ที่เกิดขึ้นจริงจากจุดที่สังเกตบนเส้นกราฟ ROC เรียกว่า “empirical ROC curve” โดยไม่มีข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงความน่าจะเป็นมาเกี่ยวข้อง ส่วนใหญ่ถูกพิจารณาภายใต้วิธีการของ Delong et al. (1988)[16] **(2) วิธีการวิเคราะห์แบบพาราเมตริก** เป็นวิธีพิจารณาพื้นที่ AUC โดยคำนึงถึงรอยเว้า (concave) และพิจารณาเส้นกราฟ ROC แบบเรียบ (smoothing) ซึ่งเรียกว่า “smoothed ROC curve” ภายใต้ข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงความน่าจะเป็นแบบปกติ 2 กลุ่ม หรือ Binormal (กลุ่มเป็นโรคและกลุ่มไม่เป็นโรค) และส่วนใหญ่ถูกพิจารณาด้วยวิธีการของ Metz (1978)[17] ดังนั้นในการนำมาใช้ จึงสามารถจำแนกตามวัตถุประสงค์ได้ดังนี้

1. การพิจารณาพื้นที่ AUC แต่ละวิธี เป็นวิธีการพิจารณาหาค่าพื้นที่ใต้โค้ง ROC หรือ AUC ในแต่ละวิธีเพื่อประเมินประสิทธิภาพของการจำแนกได้ถูกต้องโดยรวมของเครื่องมือ ซึ่งค่าผลลัพธ์ที่ได้จากโปรแกรมสำเร็จรูป จะแสดงค่าประมาณ ค่า *p*-value และช่วงเชื่อมั่นที่ระดับ 95% ของ AUC (ภายใต้สมมติฐานทางสถิติ H_0: AUC = 0.50, H_A: AUC ≠ 0.50) ขึ้นกับแนวทางในการวิเคราะห์ที่เลือกใช้ ซึ่งเป็นไปได้ทั้งแบบพาราเมตริกและแบบนอนพาราเตริก เช่น ผลลัพธ์ที่ได้จากโปรแกรม SPSS จะให้ค่าประมาณ ค่า *p*-value และช่วงเชื่อมั่นที่ระดับ 95% ของ AUC ภายใต้แนวทางแบบนอนพาราเมตริก ขณะที่โปรแกรม STATA จะให้ค่าประมาณและช่วงเชื่อมั่นที่ระดับ 95% ของ AUC ภายใต้แนวทางแบบพาราเมตริก เป็นต้น โดยมีสูตรในการพิจารณาเป็นดังนี้

$$Z = \frac{\hat{AUC} - AUC_0}{SE(\hat{AUC})}$$

เมื่อ $\hat{AUC}$ และค่าความคลาดเคลื่อนมาตรฐาน (Standard Error, SE) สามารถถูกพิจารณาได้ทั้งแบบพาราเมตริกและแบบนอนพาราเมตริก และมีเกณฑ์การตัดสินใจทั่วไป (rule of thumb) สำหรับค่าประมาณ (ดังตารางที่ 1)

ตารางที่ 1 เกณฑ์การตัดสินใจทั่วไปของค่าประมาณพื้นที่ใต้โค้ง ROC หรือ AUC[18]

ค่าพื้นที่ใต้โค้ง ROC หรือ AUC	ความหมายของการจำแนกได้อย่างถูกต้องโดยรวมของเครื่องมือ
มากกว่า 0.90	ดีมาก
0.75 – 0.90	ดี
0.50 – 0.75	ค่อนข้างต่ำ
ต่ำกว่า 0.5	ไม่สามารถนำมาใช้ได้

2. การพิจารณาเปรียบเทียบพื้นที่ AUC สองวิธี เป็นการหาค่าพื้นที่ใต้โค้ง ROC หรือ AUC เพื่อเปรียบเทียบประสิทธิภาพการจำแนกความถูกต้องโดยรวมของเครื่องมือระหว่างสองวิธี ซึ่งในทางปฏิบัติ มีความเป็นไปได้ทั้งแบบรายคู่ (คนหนึ่งคนถูกวัดด้วยเครื่องมือสองวิธี, paired sample) และแบบสองกลุ่มอิสระ (คนหนึ่งคนถูกวัดเพียงเครื่องมือเดียวจากสองวิธีที่นำมาเปรียบเทียบ, independent samples) ดังนั้นในการคำนวณค่าสถิติที่เกี่ยวข้อง จึงจำเป็นต้องคำนึงถึงความสัมพันธ์ที่เกิดขึ้นระหว่างสองวิธีที่นำมาเปรียบเทียบโดยเฉพาะในการคำนวณค่า SE โดยมีสูตรในการพิจารณาเป็นดังนี้

$$Z = \frac{\hat{AUC}_1 - \hat{AUC}_2}{SE(\hat{AUC}_1 - \hat{AUC}_2)}$$

เมื่อ

$$SE(\hat{AUC}_1 - \hat{AUC}_2) = \sqrt{Var(\hat{AUC}_1) + Var(\hat{AUC}_2) - 2Cov(\hat{AUC}_1, \hat{AUC}_2)}$$

และ $Cov(\hat{AUC}_1, \hat{AUC}_2) = r.SE(\hat{AUC}_1).SE(\hat{AUC}_2)$

เมื่อ r เป็นค่าความสัมพันธ์ระหว่างพื้นที่ใต้โค้ง (AUC) สองวิธี ซึ่งหากวิธีการทดสอบทั้งสองวิธีเป็นอิสระต่อกัน เทอมของ $Cov(\hat{AUC}_1, \hat{AUC}_2)$ ที่ได้ข้างต้น จะมีค่าเท่ากับศูนย์และหากทั้งสองวิธีมีความสัมพันธ์กัน เทอมของ $Cov(\hat{AUC}_1, \hat{AUC}_2)$ ที่ได้ จะมีค่าไม่เท่ากับศูนย์ ดังนั้นการออกแบบวิจัยตั้งแต่เริ่มต้น จึงมีผลต่อการพิจารณาความสัมพันธ์ที่เกิดขึ้นระหว่างพื้นที่ใต้โค้ง (AUC) และส่งผลโดยตรงต่อการคำนวณค่าความแปรปรวนร่วม (covariance, Cov) และค่า SE ที่ได้ ซึ่งในทางปฏิบัติวิธีการของ Delong et al. ได้ถูกนำมาใช้ในการคำนวณค่าความแปรปรวนร่วมระหว่างพื้นที่ใต้โค้ง (AUC) สองวิธีภายใต้แนวคิดแบบนอนพาราเมตริก ขณะที่วิธีการของ Metz et al. ถูกนำมาใช้คำนวณค่าความแปรปรวนร่วมระหว่างพื้นที่ใต้โค้ง (AUC) สองวิธี ภายใต้แนวคิดแบบพาราเมตริก โดยทั้งสองแนวคิดดังกล่าว สามารถพิจารณาภายใต้สมมติฐานทางสถิติเดียวกันได้ (H_0: $AUC_1 = AUC_2$, H_A: $AUC_1 \neq AUC_2$)

กรณีตัวอย่างการวิเคราะห์เส้นโค้ง ROC ด้วยโปรแกรม STATA

นักวิจัยท่านหนึ่งได้พัฒนาเครื่องมือเพื่อประเมินความเสี่ยงต่อการควบคุมระดับน้ำตาลในเลือดของผู้ป่วยเบาหวาน จำนวน 2 ชุด ภายใต้แนวคิดและทฤษฎีที่แตกต่างกัน โดยแต่ละชุดมีจำนวน 10 ข้อ และแต่ละข้อมีเกณฑ์การให้คะแนน ประกอบด้วย 0=“ไม่เสี่ยง” 1=“เสี่ยงเล็กน้อย” 2=“เสี่ยง” 3=“เสี่ยงมาก” 4=“เสี่ยงมากที่สุด” (ต่ำสุด= 0, สูงสุด= 40 คะแนน) จากนั้นนำเครื่องมือทั้งสองชุดไปประเมินผู้ป่วยเบาหวานในช่วงเดือนที่ 1 และประเมินผลระดับน้ำตาลในเลือดของแต่ละคนด้วยวิธี HbA1c ในเดือนที่ 3 เมื่อกำหนดให้ id=รหัสบุคคล, test1=คะแนนแบบประเมินด้วยเครื่องมือชุดที่ 1, test2=คะแนนแบบประเมินด้วยเครื่องมือชุดที่ 2 และ outcome=0 แทน“ควบคุมระดับน้ำตาลในเลือดได้

(HbA1c < 7.0%)" 1 แทน"ควบคุมระดับน้ำตาลในเลือดไม่ได้ (HbA1c ≥ 7.0%)" โดยมีผลการเก็บรวบรวมข้อมูลดังนี้

id	test1	test2	outcome
1	21	28	1
2	18	23	1
3	28	24	0
4	12	11	0
5	14	10	0
6	26	22	0
7	28	30	1
8	18	20	0
9	28	32	1
10	20	18	0
11	22	16	0
12	24	12	0
13	25	29	1
14	25	24	0
15	28	33	1
16	6	10	0
17	30	32	1
18	10	9	0
19	24	30	1
20	26	10	0

1. การพิจารณาหาค่าจุดตัดที่เหมาะสมของเครื่องมือนวัตกรรมทางสุขภาพชุดที่ 1 และชุดที่ 2

ในการพิจารณาหาค่าจุดตัดที่เหมาะสมทั้ง 3 วิธีข้างต้น จำเป็นต้องอาศัยค่าบ่งชี้พื้นฐานในการคำนวณ ได้แก่ ค่าความไวและความจำเพาะ ดังนั้นจึงสามารถใช้คำสั่งในโปรแกรม STATA ได้ดังตารางที่ 2

ผลลัพธ์ที่ได้จากการใช้คำสั่งในโปรแกรม STATA ดังตารางที่ 2 ตัวแปรความไวและความจำเพาะของเครื่องมือชุดที่ 1 และชุดที่ 2 จะถูกสร้างขึ้นโดยพิจารณาทุกค่าของจุดตัดที่เป็นไปได้ทั้งหมด และจากนั้นนักวิจัยสามารถนำค่าดังกล่าวไปคำนวณหาค่าจุดตัดที่เหมาะสมทั้ง 3 วิธีได้ โดยมีผลลัพธ์แสดงได้ดังตารางที่ 3

2. การพิจารณาพื้นที่ใต้โค้ง ROC หรือ AUC ของเครื่องมือนวัตกรรมทางสุขภาพชุดที่ 1 และชุดที่ 2

ในการประเมินประสิทธิภาพการจำแนกความถูกต้องในภาพรวมของเครื่องมือ โดยเฉพาะนวัตกรรมทางสุขภาพ นักวิจัยสามารถนำพื้นที่ใต้โค้ง ROC หรือ AUC ของเครื่องมือมาพิจารณาเปรียบเทียบระหว่างสองวิธีได้ ซึ่งสามารถใช้คำสั่งในโปรแกรม STATA ได้ดังตารางที่ 4

ผลลัพธ์ที่ได้จากการใช้คำสั่งในโปรแกรม STATA ดังตารางที่ 4 ข้างต้น จะได้กราฟเส้นโค้ง ROC ค่าประมาณพื้นที่ใต้โค้ง ROC หรือ AUC และช่วงเชื่อมั่นที่ระดับ 95% ของเครื่องมือชุดที่ 1 และชุดที่ 2 และกราฟแสดงการเปรียบเทียบ AUC พร้อมตารางสรุปผลการทดสอบสมมติฐานทางสถิติด้วยวิธีการของ Delong test โดยค่าสถิติทดสอบถูกแสดงในรูปค่าไคสแควร์ (χ^2) แต่ในทางทฤษฎีถือว่า เป็นค่าเดียวกับสถิติทดสอบ z^2 และมีผลลัพธ์ดังตารางที่ 5

จากผลการวิเคราะห์ดังตารางที่ 5 จึงสรุปได้ว่า "พื้นที่ใต้โค้ง ROC หรือ AUC ของเครื่องมือชุดที่ 1 และชุดที่ 2 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความผิดพลาดไม่เกิน 5% (χ^2=5.07, *p*-value=0.0243) หรืออาจกล่าวได้ว่า เครื่องมือชุดที่ 2 มีประสิทธิภาพการจำแนกความถูกต้องโดยรวมของเครื่องมือได้ดีมากกว่าเครื่องมือชุดที่ 1" อย่างไรก็ตาม ในกรณีตัวอย่างนี้ ผู้เขียนต้องการนำเสนอแนวทางการวิเคราะห์ข้อมูลเพื่อให้ครบตามประเด็นที่เกี่ยวข้อง แต่ในทางปฏิบัติจริง นักวิจัยควรให้ความสำคัญกับประเด็นระเบียบวิธีวิจัย โดยเฉพาะวิธีการมาตรฐานที่นำมาเปรียบเทียบ ควรเป็นวิธีการมาตรฐานที่เป็นที่ยอมรับ และขนาดตัวอย่างที่นำมาใช้ทำการวิเคราะห์ความมีขนาดเพียงพอ เมื่อพิจารณาจากอำนาจการทดสอบ รวมถึงการนำเสนอผลลัพธ์ที่ได้ ควรมีความครอบคลุมและสอดคล้องกับแนวทางการนำเสนอที่เป็นมาตรฐานสากล ทั้งนี้เพื่อให้ผลลัพธ์ที่ได้จากการประเมินประสิทธิภาพของเครื่องมือมีความถูกต้องและสามารถนำไปสู่การประยุกต์หรือ การขยายผลในทางปฏิบัติได้มากยิ่งขึ้น

บทสรุป

กราฟเส้นโค้ง ROC ถูกนำมาใช้ประโยชน์สำหรับการพัฒนาเครื่องมือที่เป็นนวัตกรรมทางสุขภาพ โดยเฉพาะกรณีเครื่องมือที่มีผลลัพธ์แบบต่อเนื่อง หรือ อันดับ ใน 2 ประเด็นหลัก ได้แก่ (1) การพิจารณาค่าจุดตัดที่เหมาะสม ประกอบด้วย วิธี Euclidean index วิธี Youden's index และวิธี Weighted Number Need to Misdiagnose (Weighted NNM) และ (2) การประเมินประสิทธิภาพการจำแนกความถูกต้องโดยรวมของเครื่องมือด้วยการวิเคราะห์พื้นที่ใต้โค้ง ROC หรือ AUC ประกอบด้วย การวิเคราะห์แบบนอนพาราเมตริก ภายใต้การไม่มีข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงความน่าจะเป็นและมีการพิจารณาเส้นโค้งแบบ empirical ROC curve และการวิเคราะห์แบบพาราเมตริก ซึ่งมีข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงความน่าจะเป็นแบบปกติสองกลุ่ม หรือ Binormal และพิจารณาเส้นโค้งแบบ smoothed ROC curve โดยทั้งสองวิธีดังกล่าว สามารถนำมาใช้ทำการวิเคราะห์พื้นที่ใต้โค้ง ROC หรือ AUC ได้ทั้งแบบวิธีเดียว และแบบสองวิธี ภายใต้ความพร้อมของโปรแกรมสำเร็จรูปทางสถิติในการรองรับการวิเคราะห์อย่างครอบคลุม อย่างไรก็ตาม ในทางปฏิบัติวิธีการมาตรฐานที่นำมาเปรียบเทียบ ขนาดตัวอย่างที่ใช้ และการนำเสนอผลลัพธ์ที่ครอบคลุม ยังเป็นประเด็นที่นักวิจัยควรให้ความสำคัญ เพื่อให้ผลลัพธ์ที่ได้จากการประเมินประสิทธิภาพของเครื่องมือมีความถูกต้องและสามารถนำไปสู่การประยุกต์ หรือ การขยายผลในทางปฏิบัติได้มากยิ่งขึ้น

ตารางที่ 2 คำสั่งสำหรับการคำนวณค่าความไว, ความจำเพาะ และจุดตัดที่เหมาะสมของเครื่องมือ

กรณีเครื่องมือชุดที่ 1 (test1)	**กรณีเครื่องมือชุดที่ 2 (test2)**
การคำนวณค่าความไวและความจำเพาะ	
`quietly logit outcome test1` `lsens, gensens(sen1) genspec(spec1)`	`quietly logit outcome test2` `lsens, gensens(sen2) genspec(spec2)`
การคำนวณหาจุดตัดที่เหมาะสมด้วยวิธี Euclidian's index (i)	

`gen float i1=sqrt((1-sen1)^2+(1-spec1)^2)`
`egen float min1=min(i1)`
`list test1 if i1==min1`
	`gen float i2=sqrt((1-sen2)^2+(1-spec2)^2)`
	`egen float min2=min(i2)`
	`list test2 if i2==min2`

การคำนวณหาจุดตัดที่เหมาะสมด้วยวิธี Youden's index (J)	

`gen float youden1=sen1+spec1-1`
`egen float max1=max(youden1)`
`list test1 if youden1==max1`
	`gen float youden2=sen2+spec2-1`
	`egen float max2=max(youden2)`
	`list test2 if youden2==max2`

การคำนวณหาจุดตัดที่เหมาะสมด้วยวิธี Weighted NNM (เมื่อกำหนดให้ C=3, pr=0.3)	

`gen float wnnm1=1/(3*(0.3-sen1))+(0.7*(1-spec1))`
`egen float wmax1=max(wnnm1)`
`list test1 if wnnm1==wmax1`
	`gen float wnnm2=1/(3*(0.3-sen2))+(0.7*(1-spec2))`
	`egen float wmax2=max(wnnm2)`
	`list test2 if wnnm2==wmax2`

ตารางที่ 3 ค่าจุดตัดคะแนนที่เหมาะสม จำแนกตามวิธีที่ถูกนำมาพิจารณา

วิธีการพิจารณา	**เครื่องมือชุดที่ 1 (test1)**	**เครื่องมือชุดที่ 2 (test2)**
	ค่าจุดตัดคะแนนที่เหมาะสม	**ค่าจุดตัดคะแนนที่เหมาะสม**
วิธี Euclidian's index (i)	22	24
วิธี Youden's index (J)	26	24
วิธี Weighted NNM (C=3, pr=0.3)	28	32

ตารางที่ 4 คำสั่งสำหรับการคำนวณค่าความไว, ความจำเพาะ และจุดตัดที่เหมาะสมของเครื่องมือ

กรณีเครื่องมือชุดที่ 1 (test1)	**กรณีเครื่องมือชุดที่ 2 (test2)**
การสร้างกราฟเส้นโค้ง ROC	
`quietly logit outcome test1` `lroc`	`quietly logit outcome test2` `lroc`
การคำนวณหาพื้นที่ใต้เส้นโค้ง ROC หรือ AUC	
`roctab outcome test1`	`roctab outcome test2`
การสร้างตัวแปรทำนาย	
`predict ptest1,xb`	`predict ptest2,xb`
การเปรียบเทียบพื้นที่ใต้โค้ง ROC หรือ AUC	
`roccomp outcome ptest1 ptest2,graph summary`	

ตารางที่ 5 ผลการวิเคราะห์พื้นที่ใต้โค้ง ROC หรือ AUC ของเครื่องมือสองวิธี

เครื่องมือ	**ตัวอย่าง (n)**	**AUC**	**95%CI of AUC**	$\chi^2 \approx Z^2$	***p*-value**
ชุดที่ 1	20	0.7604	0.5428 – 0.9780	5.07	0.0243
ชุดที่ 2	20	0.9792	0.9299 - 1.0000		

ข้อเสนอแนะเพื่อการนำไปใช้

ในการวิเคราะห์เส้นโค้ง ROC เพื่อนำมาใช้ในการประเมินจุดตัดที่เหมาะสมและประสิทธิภาพการจำแนกความถูกต้องโดยรวมของเครื่องมือที่เป็นนวัตกรรมทางสุขภาพ มีข้อเสนอแนะเพื่อการนำไปใช้ ซึ่งเป็นทั้งจุดเด่นและข้อจำกัดที่พึงระวัง จำแนกเป็นรายข้อได้ดังนี้

1. ผลลัพธ์ที่ได้จากการวิเคราะห์เส้นโค้ง ROC ไม่ขึ้นกับค่าสเกลการวัดของเครื่องมือและไม่ได้รับผลกระทบโดยตรงจากอัตราความชุกของโรค แต่อย่างไรก็ตาม ในการพิจารณาเพื่อนำเครื่องมือไปใช้ในทางปฏิบัติ นักวิจัยอาจจำเป็นต้องคำนึงถึงคุณลักษณะทางประชากรของผู้ป่วยหรือ กลุ่มเป้าหมายที่นำไปใช้ รวมถึงขอบเขตอาการแสดงของโรค (disease spectrum)[19] ที่เป็นผลลัพธ์หลักในการศึกษาด้วย

2. การพิจารณาวิธีการวิเคราะห์เส้นโค้ง ROC ส่วนใหญ่ รวมถึงบทความครั้งนี้ เป็นการพิจารณาสถานะการ

เกิดผลลัพธ์จริงและผลลัพธ์ที่ได้จากการประเมินด้วยเครื่องมือภายใต้เวลาที่กำหนดให้คงที่ (fixed over time) ซึ่งอาจเป็นข้อจำกัดที่นักวิจัยควรพึงระวัง เพราะในทางปฏิบัติ ผลลัพธ์ทั้งสองดังกล่าว มีความผันแปรตามเวลาโดยตลอด (change over time)[20]

3. ขนาดตัวอย่างที่มีจำนวนน้อย อาจส่งผลกระทบโดยตรงต่อความคลาดเคลื่อนของค่าประมาณที่ได้ ดังนั้น วิธี Re-Sampling ด้วยวิธี Jackknife และ bootstrapping จึงอาจเป็นทางเลือกที่นักวิจัยควรพิจารณานำมาใช้ เช่น การวิเคราะห์ด้วยคำสั่ง comproc ในโปรแกรม STATA เป็นต้น[21]

4. การวิเคราะห์ AUC แบบนอนพาราเมตริก กรณีเปรียบเทียบเครื่องมือสองวิธี เมื่อข้อมูลถูกพิจารณาแบบรายคู่ (คนหนึ่งถูกวัดสองวิธี) และแบบอิสระ (แต่ละวิธีวัดแยกคนแต่ละกลุ่ม) การนำวิธีการ Delong test มาใช้ในการวิเคราะห์ถือว่า เหมาะสม แต่ในกรณีข้อมูลสองวิธีที่นำมาเปรียบเทียบ มีลักษณะแบบซ้อนทับ (nested) การใช้วิธีการ Delong test อาจมีข้อจำกัดและไม่เหมาะสม[22]

เอกสารอ้างอิง

1. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb). 2016;26(3):297-307.
2. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Diagnostic methods 2: receiver operating characteristic (ROC) curves. Kidney Int. 2009;76(3):252-6.
3. Ren Q, Su C, Wang H, Wang Z, Du W, Zhang B. Prospective study of optimal obesity index cut-off values for predicting incidence of hypertension in 18-65-year-old Chinese adults. PloS one. 2016;11(3):e0148140-e0148140.
4. Silva PAB, Soares SM, Santos JFG, Silva LB. Cut-off point for WHOQOL-bref as a measure of quality of life of older adults. Rev Saude Publica. 2014;48(3):390-7.
5. Westergren A, Norberg E, Vallén C, Hagell P. Cut-off scores for the Minimal Eating Observation and Nutrition Form – Version II (MEONF-II) among hospital inpatients. Food Nutr Res. 2011;55(1):7289.
6. Sung YH, Cho MS, Kwon IG, Jung YY, Song MR, Kim K, et al. Evaluation of falls by inpatients in an acute care hospital in Korea using the Morse Fall Scale. Int J Nurs Pract. 2014;20(5):510-7.
7. Kim HR, Kim HS. Optimal cutoffs of cardiometabolic risk for postmenopausal Korean women. Asian Nurs Res (Korean Soc Nurs Sci). 2017;11(2):107-12.
8. Chan KTM, Cheung AHN. Application of receiver operating characteristic (ROC) curve to determine the diagnostic ability of a validated ten-item questionnaire (SS - 10) in estimating the prevalence of sensitive skin in Hong Kong population. Int J Inn Res Med Sci. 2019;4(07):405-13.
9. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19(1):64.
10. Kumar R. Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr. 2011;48(4):277-87.
11. Zou KH, O' Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007;115(5):654-7.
12. Hajian-Tilaki K. The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. Stat Methods Med Res. 2018;27(8):2374-83.
13. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol. 2006;163(7):670-5.
14. Habibzadeh F, Habibzadeh P, Yadollahie M. Criterion used for determination of test cut-off value. Diabetes Res Clin Pract. 2017;128:138-9.
15. Habibzadeh F, Yadollahie M. Number needed to misdiagnose: a measure of diagnostic test effectiveness. Epidemiology. 2013;24(1):170.
16. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
17. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8(4):283-98.
18. Ray P, Manach YL, Riou B, Houle TT. Statistical Evaluation of a Biomarker. Anesthesiology. 2010;112(4):1023-40.

19. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010;5(9):1315-16.
20. Kamarudin AN, Cox T, R Kolamunnage- Dona. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Med Res Methodol. 2017;17(1):53.
21. Pepe M, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. Stata J. 2009;9(1):1.
22. Demler OV, Pencina MJ, D'Agostino RB Sr. Misuse of DeLong test to compare AUCs for nested models. Stat Med. 2012;31(23):2577-87.