



ความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกผู้ป่วยโรคโควิด-19 ของรังสีแพทย์ที่มีความต่างประสบการณ์

วิชุนันท์ กันชัย

กลุ่มงานรังสีวิทยา โรงพยาบาลชัยภูมิ จังหวัดชัยภูมิ

Inter-observer Reliability for Chest X-Ray Interpretation in COVID-19 Patients of Radiologists with Different Experiences

Wichunun Kunchai

Division of Radiology, Chaiyaphum Hospital, Chaiyaphum Province

Received: 19 October 2023 /Review: 20 October 2023 / Revised: 21 January 2024 /

Accepted: 29 January 2024

บทคัดย่อ

หลักการและวัตถุประสงค์: ภาพถ่ายทางรังสีมีบทบาทสำคัญมากในการวินิจฉัยและวางแผนการดูแลรักษาผู้ป่วย ดังนั้นการศึกษา
นี้จึงมีวัตถุประสงค์เพื่อศึกษาความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยติดเชื้อโรคโควิด-19 ของรังสีแพทย์
ที่มีประสบการณ์และรังสีแพทย์ใหม่

วิธีการศึกษา: เป็นการศึกษาย้อนหลังเชิงพรรณนา ในผู้ป่วยติดเชื้อโรคโควิด-19 ของโรงพยาบาลชัยภูมิระหว่างเดือนสิงหาคม พ.ศ.
2564 ถึงกุมภาพันธ์ พ.ศ. 2565 แปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยโรคโควิด-19 โดยรังสีแพทย์ 3 ราย ประสบการณ์ 11 ปี 9
ปี และรังสีแพทย์ใหม่ โดยใช้ Brixia scoring system ซึ่งแบ่งปอดในภาพถ่ายรังสีทรวงอกออกเป็น 6 โซน รังสีแพทย์ 1 ราย ทำการ
แปลผลภาพถ่ายรังสี 2 ครั้งต่างช่วงเวลาเพื่อประเมินความเที่ยงของผู้ประเมิน วิเคราะห์ความสอดคล้องระหว่างผู้ประเมินใช้สถิติแคป
ปาและค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น ทั้งโดยภาพรวมและเฉพาะโซนของภาพถ่ายรังสี ค่าความเชื่อมั่นที่ร้อยละ 95

ผลการศึกษา: การศึกษานี้พบว่าความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยโรคโควิด-19 ของรังสีแพทย์ 3 ราย
มีความสอดคล้องกันโดยรวมในระดับดีถึงดีมาก (weighted Cohen's $K = 0.79-0.90$) ความสอดคล้องในการแปลผลระหว่าง
รังสีแพทย์ที่มีประสบการณ์ดีกว่ารังสีแพทย์ใหม่เล็กน้อย ค่าความน่าเชื่อถือภายในผู้ประเมินทั้ง 3 (intra-rater reliability) โดยรวม
สอดคล้องกันในระดับดี (ICC 0.88: 95%CI = 0.84-0.91) ส่วนความเที่ยงของผู้ประเมิน (intra-observer reliability) พบว่า มีความ
น่าเชื่อถือสอดคล้องกันในระดับดีมาก (ICC 0.97: 95%CI = 0.95-0.98 และ Pearson correlation =0.97, $p<0.001$)

สรุป: ความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยโรคโควิด-19 ของรังสีแพทย์ 3 ราย ทั้งรังสีแพทย์ที่มีประสบการณ์
และรังสีแพทย์ใหม่ มีความสอดคล้องกันโดยรวมในระดับดีถึงดีมาก

คำสำคัญ: ความสอดคล้องระหว่างผู้ประเมิน, ภาพถ่ายรังสีทรวงอกโรคโควิด-19, ประเมินภาวะปอดอักเสบจากโรคโควิด-19

*Corresponding author: Wichunun Kunchai, Email: wichununk@gmail.com

Abstract

Background and Objectives: Radiographs play a very important role in diagnosing and planning patient care. This study aimed to assess the inter- and intraobserver reliability among radiologists, particularly between new and experienced radiologists, in evaluating the extent of pulmonary abnormalities in COVID-19 patients on chest X-rays (CXRs).

Methodology: Cross-sectional study used retrospective data from August 2021 to February 2022, 119 chest radiographs from PCR confirmed COVID-19 patients in Chaiyaphum hospital were reviewed by 3 radiologists. Three radiologists, with 11 and 9 years of experience, as well as one new radiologist, independently reviewed all chest X-rays using the Brixia scoring system, which divides the CXR into six lung zones. One radiologist reviewed all CXRs twice to evaluate intra-observer reliability. The inter-observer agreement was use weighted Cohen's Kappa, the Intra-class correlation coefficients in the overall and zone-specific were calculated and along with 95% confidence interval (95%CI).

Results: This study found the overall inter-observer agreement between radiologist show substantial to almost perfect (weighted Cohen's $K = 0.79-0.90$). Concordance in interpretation between experienced radiologists was slightly better than concordance in interpretation between experienced and new radiologists. Intra-class correlation coefficients (ICC) in the overall show good correlation (ICC 0.88:95%CI = 0.84-0.91). The intra-observer reliability of one experienced radiologist shows very good correlation (ICC 0.97: 95%CI = 0.95-0.98 and Pearson correlation coefficient = 0.97, $p < 0.001$).

Conclusions: Substantial to almost perfect interobserver agreement between experience and new radiologist in COVID-19 chest radiograph interpretation.

Keywords: interobserver reliability, COVID-19 CXR, COVID pneumonia severity score

บทนำ

โรคติดเชื้อไวรัสโคโรนาสายพันธุ์ใหม่หรือโรคโควิด-19 เป็นโรคติดเชื้อระบบทางเดินหายใจที่เกิดจากเชื้อโคโรนาไวรัส 2 (SARS-CoV-2) องค์การอนามัยโลกได้ประกาศให้โรคติดเชื้อโควิด-19 ถือเป็นโรคระบาดไปทั่วโลกเมื่อวันที่ 11 มีนาคม พ.ศ. 2563^{1,2}

ภาพถ่ายทางรังสีมีบทบาทสำคัญมากในการวินิจฉัยและวางแผนการดูแลรักษาผู้ป่วยโรคโควิด-19 โดยภาพถ่ายรังสีทรวงอก (chest X-ray) เป็นการตรวจวินิจฉัยทางรังสีที่ใช้เป็นอันดับแรกและเป็นหลักในการประเมินความผิดปกติของปอดจากการติดเชื้อโรคโควิด-19 ในโรงพยาบาลชัยภูมิ มีประโยชน์ในการวินิจฉัยภาวะปอดอักเสบ รวมถึงติดตามประเมินผลการรักษา แม้ว่าจะมีความไวในการประเมินความผิดปกติของปอดในระยะแรกน้อยกว่าการตรวจเอกซเรย์คอมพิวเตอร์ (CT scan of chest)³⁻⁵ แต่ภาพถ่ายรังสีทรวงอกเป็นการตรวจวินิจฉัยที่เครื่องมือมีความแพร่หลายเข้าถึงได้มากกว่า รวมถึงสะดวกและรวดเร็วกว่า อย่างไรก็ตามระยะของการติดเชื้อ สภาวะปัจจุบันของผู้ป่วย และพยาธิสภาพเดิมของปอดอาจส่งผลให้เกิดความผิดพลาดในการประเมินภาพถ่ายรังสีทรวงอก

การวินิจฉัยภาวะปอดอักเสบจากการติดเชื้อโรคโควิด-19 ได้อย่างรวดเร็วมีความสำคัญมากในการวางแผนการรักษาผู้ป่วย โดยเฉพาะอย่างยิ่งในภาวะที่ยาต้านไวรัสที่ใช้ในการรักษามีจำกัด ดังนั้นการวินิจฉัยภาวะปอดอักเสบในผู้ป่วยโรคโควิด-19 จากภาพถ่ายรังสีทรวงอกที่ถูกต้องมีความจำเป็นอย่างยิ่งต่อการวางแผนการรักษาที่มีประสิทธิภาพ

การประเมินภาวะปอดอักเสบในผู้ป่วยโรคโควิด-19 จากภาพถ่ายรังสีทรวงอก สามารถทำได้หลายวิธี อาทิ CXR score (Brixia score)⁶, SARI CXR severity scoring system⁷, และ radiographic assessment of lung edema (RALE score)⁸ โดย Brixia scoring system เป็นการประเมินภาวะปอดอักเสบจากเชื้อโควิด-19 ที่ใช้อ้างอิงแพร่หลายที่สุด

ข้อมูลการศึกษาความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยติดเชื้อโรคโควิด-19 (inter- and intraobserver variability) ระหว่างรังสีแพทย์ยังมีจำกัด โดยเฉพาะอย่างยิ่งระหว่างรังสีแพทย์ที่มีประสบการณ์และรังสีแพทย์ใหม่ ในการประเมินความผิดปกติของปอดจากภาพถ่ายรังสีทรวงอกโดยใช้ Brixia score ซึ่งข้อมูลที่ยกมามีผลต่อการวางแผนการรักษาผู้ป่วย ดังนั้นวัตถุประสงค์ของการศึกษานี้ทำขึ้นเพื่อศึกษาความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยติดเชื้อโรคโควิด-19 ระหว่างรังสีแพทย์ โดยเฉพาะอย่างยิ่งระหว่างรังสีแพทย์ที่มีประสบการณ์และรังสีแพทย์ใหม่โดยใช้ Brixia scoring system

วิธีการศึกษา

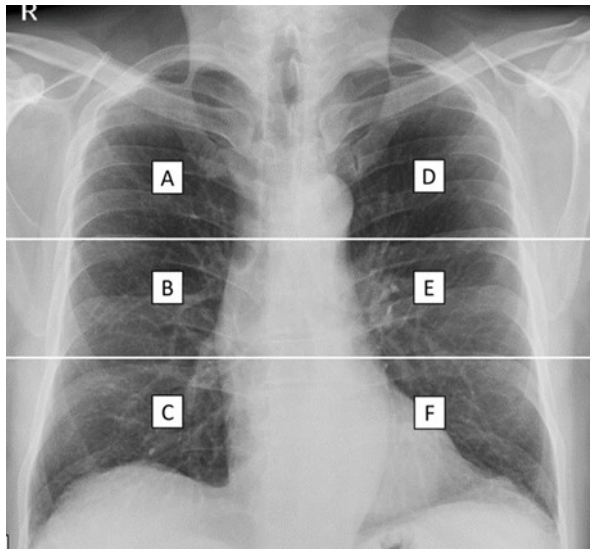
การศึกษานี้ได้ผ่านการรับรองจากคณะกรรมการจริยธรรมการวิจัยเกี่ยวกับมนุษย์ โรงพยาบาลชัยภูมิ เลขที่ 035/2566 เป็นการศึกษาย้อนหลังเชิงพรรณนา ในผู้ป่วยติดเชื้อโรคโควิด-19 ที่ทำการรักษาในโรงพยาบาลชัยภูมิ ระหว่างเดือนสิงหาคม พ.ศ. 2564 ถึงกุมภาพันธ์ พ.ศ. 2565 สืบค้นจากระบบสารสนเทศของโรงพยาบาล เกณฑ์การคัดเข้าศึกษา (inclusion criteria) ได้แก่ ผู้ป่วยติดเชื้อโรคโควิด-19 ที่วินิจฉัยโดยการตรวจพบสารพันธุกรรมของเชื้อไวรัสจากทางเดินหายใจส่วนบน (RT-PCR positive for SAR-CoV-2) และมีภาพถ่ายรังสีทรวงอกในระบบเก็บภาพถ่ายรังสีของโรงพยาบาล (Picture archiving and communication system :PACS) เกณฑ์การคัดออกจากการศึกษา (exclusion criteria) ได้แก่ ข้อมูลภาพถ่ายทางรังสีเป็นภาพถ่ายที่ไม่สามารถแปลผลได้

การคำนวณขนาดตัวอย่าง

ในการศึกษาก่อนหน้าที่ดำเนินการโดย Monaco และคณะ⁹ พบว่า ค่า intraclass correlation coefficients (ICC) โดยรวมคือ 0.7 สำหรับการศึกษาที่ใช้ ICC เช่นเดียวกับการศึกษาก่อนหน้านี้ ดังนั้น ICC มีค่าเท่ากับ 0.7 นอกจากนี้ มีการใช้ระดับ alpha ที่เท่ากับ 0.05 พร้อมกับมีผู้ประเมิน 3 ราย และความกว้างของความน่าเชื่อถือของค่าหลัก 0.15 เพื่อคำนวณขนาดตัวอย่าง ต้องใช้ขนาดตัวอย่างประมาณ 119 ตัวอย่างในการประเมิน (คำนวณโดยใช้ ส่วนบนของฟอร์ม The STATA software version 18.0 by “sampicc” command)¹⁰

การศึกษานี้ใช้ภาพถ่ายรังสีทั้งหมด 119 ภาพ จากผู้ป่วยโรคโควิด-19 จำนวน 92 ราย เก็บรวบรวมข้อมูลผู้ป่วยจากระบบสารสนเทศของโรงพยาบาล ประเมินภาพถ่ายรังสีทรวงอก (chest x-ray) จากระบบเก็บภาพถ่ายทางรังสี (PACS) โดยรังสีแพทย์ 3 ราย แบ่งเป็นรังสีแพทย์ประสบการณ์ 11 ปี 9 ปี และรังสีแพทย์ใหม่ประสบการณ์ทำงาน 3 เดือน แยกกันแปลผลภาพถ่ายทางรังสีแบบปกปิดข้อมูลผู้ป่วย

ประเมินภาพถ่ายรังสีทรวงอกโดยใช้ Brixia score system⁶ ซึ่งจะแบ่งพื้นที่ของปอดออกเป็น 6 โซน ได้แก่ โซนบน (upper zones: ขวา A และซ้าย D), โซนกลาง (middle zones: ขวา B และซ้าย E) และโซนล่าง (lower zones: ขวา C และซ้าย F) (รูปที่ 1) โดยระบุความผิดปกติที่พบในปอดในแต่ละส่วนเป็นคะแนน ดังนี้ no abnormality = 0 คะแนน, interstitial infiltration = 1 คะแนน, interstitial and alveolar infiltrates (interstitial predominance) = 2 คะแนน และ interstitial and alveolar infiltrates (alveolar predominance) = 3 คะแนน โดยคะแนนความผิดปกติของภาพถ่ายรังสี (CXR score) จะมีค่าตั้งแต่ 0-18



รูปที่ 1 แบ่งพื้นที่ของปอดในภาพถ่ายรังสีทรวงอกเป็น 6 โซน โดยเส้นบนอยู่ที่ระดับขอบล่างของ aortic arch เส้นล่างอยู่ที่ระดับขอบล่างของ right inferior pulmonary vein

การประเมินความน่าเชื่อถือภายในผู้ประเมิน (Intra-observer reliability) ทำโดย รังสีแพทย์ 1 ราย ทำการแปลผลภาพถ่ายทางรังสีทรวงอกทั้งหมด 2 ครั้ง ในระยะเวลาห่างกันเป็นเวลา 6 สัปดาห์ เพื่อลดอคติในการประเมิน

การวิเคราะห์ทางสถิติ

ข้อมูลอายุผู้ป่วยรายงานด้วยค่าสถิติเชิงพรรณนา (ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ข้อมูลเชิงกลุ่มรายงานเป็นความถี่และร้อยละ) ความสอดคล้องระหว่างผู้ประเมินใช้สถิติ Kappa (weighted Cohen's Kappa) และค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (ICC) ทั้งโดยภาพรวมและเฉพาะโซนของภาพถ่ายรังสี ค่าความเชื่อมั่นที่ร้อยละ 95 (95% confidence interval: CI) วิเคราะห์ข้อมูลโดย STATA version 18

การแปลความหมายขนาดความสอดคล้อง (strange of agreement) ของค่าสถิติ Kappa พิจารณาดังนี้ : ค่า Kappa <0.00 แย่ (poor), 0.00-0.20 น้อย (slight), 0.21-0.40 พอใช้ (fair), 0.41-0.60 ปานกลาง (moderate), 0.61-0.80 ดี (substantial) และ 0.81-1.00 ดีมาก/ค่อนข้างสมบูรณ์ (almost perfected)¹¹

การวิเคราะห์ค่า ICC ซึ่งมีค่าอยู่ระหว่าง 0-1 สามารถแปลผลได้ดังนี้ : ค่า ICC 0.00-0.50 – ความน่าเชื่อถือของผู้ประเมินมีความสอดคล้องกันในระดับต่ำ, ค่า ICC 0.51-0.74 – สอดคล้องกันในระดับพอใช้, ค่า ICC 0.75-0.90 – สอดคล้องกันในระดับดี และ ค่า ICC 0.91-1.00 – สอดคล้องกันในระดับดีมาก¹²

ผลการศึกษา

อายุเฉลี่ยของผู้ป่วยโรคโควิด-19 ในการศึกษาี้เท่ากับ 50.0 ± 20.2 ปี เป็นเพศชายและหญิงในสัดส่วนที่เท่ากัน ร้อยละ 39.2 ของประชากรศึกษามีโรคประจำตัว โดยโรคประจำตัวที่พบมากที่สุดคือ โรคความดันโลหิตสูง ร้อยละ 22.8 (ตารางที่ 1)

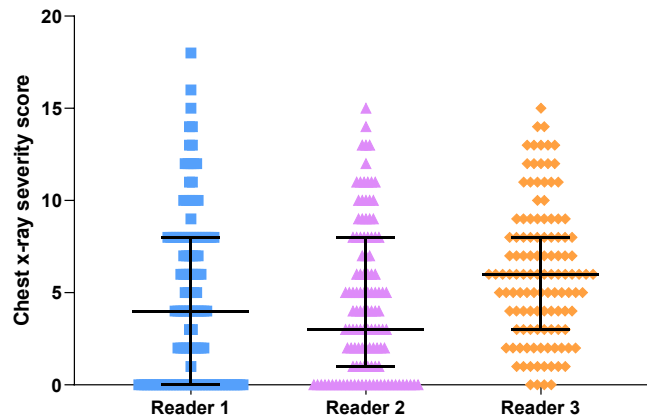
ตารางที่ 1 ลักษณะเฉพาะของผู้ป่วย

ลักษณะเฉพาะของผู้ป่วย	N=92 จำนวน (ร้อยละ)
อายุ (ปี), mean $\pm$ SD	50.0 $\pm$ 20.2
เพศ	
ชาย	46 (50.0)
หญิง	46 (50.0)
โรคประจำตัว	
โรคเบาหวาน	9 (9.8)
โรคความดันโลหิตสูง	21 (22.8)
โรคไตเรื้อรัง	3 (3.3)
โรคหอบหืด	3 (3.3)

จากตารางที่ 2 และรูปที่ 2 ค่ามัธยฐานของคะแนนรวม chest x-ray severity score ของรังสีแพทย์ทั้ง 3 ราย พบว่า รังสีแพทย์ประสบการณ์ 11 ปี (reader 1) มีค่ามัธยฐานของคะแนนรวมเท่ากับ 4 (IQR 0-8), รังสีแพทย์ประสบการณ์ 9 ปี (reader 2) มีค่ามัธยฐานของคะแนนรวมเท่ากับ 3 (IQR 1-8) และรังสีแพทย์ใหม่ (reader 3) มีค่ามัธยฐานของคะแนนรวมเท่ากับ 6 (IQR 3-8)

ตารางที่ 2 สรุปคะแนนความรุนแรงของการถ่ายภาพรังสีทรวงอกในผู้ป่วยโรคโควิด-19

	รังสีแพทย์ประสบการณ์ 11 ปี (reader 1) จำนวน (ร้อยละ)	รังสีแพทย์ประสบการณ์ 9 ปี (reader 2) จำนวน (ร้อยละ)	รังสีแพทย์ใหม่ (reader 3) จำนวน (ร้อยละ)
รวม, ค่ามัธยฐาน (IQR)	4 (0-8)	3 (1-8)	6 (3-8)
โซน A			
- 0	89 (74.8)	84 (70.6)	51 (42.9)
- 1	14 (11.8)	19 (16)	53 (44.5)
- 2	14 (11.8)	16 (13.5)	14 (11.8)
- 3	2 (1.7)	0	1 (0.8)
โซน B			
- 0	66 (55.9)	66 (55.5)	28 (23.5)
- 1	11 (9.3)	23 (19.3)	56 (47.1)
- 2	35 (29.7)	27 (22.7)	32 (26.9)
- 3	6 (5.1)	3 (2.5)	3 (2.5)
โซน C			
- 0	52 (43.7)	49 (41.2)	16 (13.5)
- 1	11 (9.2)	30 (25.2)	53 (44.5)
- 2	48 (40.3)	36 (30.3)	47 (39.5)
- 3	8 (6.7)	4 (3.4)	3 (2.5)
โซน D			
- 0	95 (79.8)	92 (77.3)	54 (45.4)
- 1	9 (7.6)	20 (16.8)	53 (44.5)
- 2	14 (11.8)	7 (5.9)	12 (10.1)
- 3	1 (0.8)	0	0
โซน E			
- 0	60 (50.4)	54 (45.4)	32 (26.9)
- 1	6 (5)	35 (29.4)	57 (47.9)
- 2	44 (37)	26 (21.9)	26 (21.9)
- 3	9 (7.6)	4 (3.4)	4 (3.4)
โซน F			
- 0	47 (39.5)	36 (30.3)	19 (16)
- 1	7 (5.9)	37 (31.1)	46 (38.7)
- 2	60 (50.4)	39 (32.8)	43 (36.1)
- 3	5 (4.2)	7 (5.9)	11 (9.2)



รูปที่ 2 คะแนนความรุนแรงของภาพถ่ายรังสีทรวงอกทั้งหมดโดยรังสีแพทย์ 3 ราย มีนัยสำคัญทางสถิติ ($p < 0.001$)

ตารางที่ 3 ค่า K ของข้อตกลงระหว่าง ที่ถ่วงน้ำหนักของ Cohen โดยรังสีแพทย์ทั้ง 2 ราย

	Reader 1	Reader 2
รวม		
- Reader 2	0.90 (0.85-0.93)	-
- Reader 3	0.82 (0.81-0.86)	0.79 (0.77-0.80)
โซน A		
- Reader 2	0.75 (0.61-0.88)	-
- Reader 3	0.64 (0.51-0.76)	0.65 (0.53-0.76)
โซน B		
- Reader 2	0.84 (0.77-0.91)	-
- Reader 3	0.72 (0.64-0.80)	0.71 (0.62-0.80)
โซน C		
- Reader 2	0.83 (0.76-0.90)	-
- Reader 3	0.72 (0.65-0.80)	0.67 (0.58-0.77)
โซน D		
- Reader 2	0.69 (0.52-0.85)	-
- Reader 3	0.60 (0.46-0.74)	0.55 (0.49-0.60)
โซน E		
- Reader 2	0.78 (0.70-0.86)	-
- Reader 3	0.72 (0.65-0.80)	0.80 (0.74-0.86)
โซน F		
- Reader 2	0.72 (0.61-0.82)	-
- Reader 3	0.68 (0.58-0.78)	0.70 (0.60-0.80)

ตารางที่ 3 ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมิน ระหว่าง reader 1 เทียบกับ reader 2 และ reader 3 ผลการอ่านคะแนนโดยรวมมีความสอดคล้องกันในระดับดีมาก ค่า weighted Cohen's **K** เท่ากับ 0.90 และ 0.82 ตามลำดับ ส่วนความสอดคล้องของ reader 2 และ reader 3 อยู่ในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.79

เมื่อพิจารณาในแต่ละโซน พบว่า

โซน A: ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมินระหว่าง reader 1 เทียบกับ reader 2 และ reader 3 ผลการอ่านคะแนนโดยรวมมีความสอดคล้องกันในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.75 และ 0.64 ตามลำดับ ความสอดคล้องของ reader 2 และ reader 3 อยู่ในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.65

โซน B: ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมินระหว่าง reader 1 เทียบกับ reader 2 มีความสอดคล้องกันในระดับดีมาก ค่า weighted Cohen's **K** เท่ากับ 0.84 และ reader 3 มีความสอดคล้องกันในระดับดี weighted Cohen's **K** เท่ากับ 0.72 และความสอดคล้องของ reader 2 และรังสีแพทย์ใหม่ (reader 3) อยู่ในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.71

โซน C: ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมินระหว่าง reader 1 เทียบกับ reader 2 มีความสอดคล้องกันในระดับดีมาก ค่า weighted Cohen's **K** เท่ากับ 0.83 และ reader 3 มีความสอดคล้องกันในระดับดี weighted Cohen's **K** เท่ากับ 0.72 และความสอดคล้องของ reader 2 และ reader 3 อยู่ในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.67

โซน D: ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมินระหว่าง reader 1 เทียบกับ reader 2 มีความสอดคล้องกันในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.69 และ reader 3 มีความสอดคล้องกันในระดับปานกลาง weighted Cohen's **K** เท่ากับ 0.60 และความสอดคล้องของ reader 2 และ reader 3 อยู่ในระดับปานกลาง ค่า weighted Cohen's **K** เท่ากับ 0.55

โซน E: ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมินระหว่าง reader 1 เทียบกับ reader 2 และ reader 3 ผลการอ่านคะแนนโดยรวมมีความสอดคล้องกันในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.78, 0.72 ตามลำดับ และความสอดคล้องของ reader 2 และ reader 3 อยู่ในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.80

โซน F: ค่าความสอดคล้องในการอ่านผลภาพถ่ายรังสีของผู้ประเมินระหว่าง reader 1 เทียบกับ reader 2 และ reader 3 ผลการอ่านคะแนนโดยรวมมีความสอดคล้องกันในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.72, 0.68 ตามลำดับ และความสอดคล้องของ reader 2 และ reader 3 อยู่ในระดับดี ค่า weighted Cohen's **K** เท่ากับ 0.70

ตารางที่ 4 ค่าสัมประสิทธิ์สหสัมพันธ์ภายในคลาสโดยรวมและเฉพาะโซน

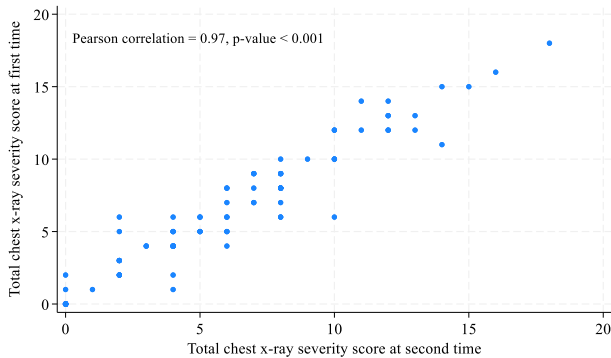
โซน	ICC (95% CI)	p-value
โดยรวม	0.88 (0.84-0.91)	<0.001
A	0.70 (0.61-0.91)	<0.001
B	0.78 (0.72-0.84)	<0.001
C	0.78 (0.70-0.82)	<0.001
D	0.77 (0.70-0.82)	<0.001
E	0.77 (0.70-0.82)	<0.001
F	0.72 (0.64-0.78)	<0.001

จากตารางที่ 4 ค่าความน่าเชื่อถือภายในผู้ประเมิน (intra-rater reliability) ของรังสีแพทย์ทั้ง 3 ราย โดยรวมเท่ากับ 0.88 ซึ่งถือว่าความสอดคล้องกันในระดับดี เมื่อพิจารณาในแต่ละโซนค่า ค่าความน่าเชื่อถือภายในผู้ประเมิน (intra-rater Reliability) ทั้ง 3 ผู้ประเมิน พบว่าค่าความน่าเชื่อถืออยู่ในระดับพอใช้ ได้แก่ โซน A และ F ค่าความน่าเชื่อถือเท่ากับ 0.70 และ 0.72 ตามลำดับ ค่าความน่าเชื่อถืออยู่ในระดับดี ได้แก่ โซน B, C, D และ E ค่าความน่าเชื่อถือเท่ากับ 0.78, 0.78, 0.77 และ 0.77 ตามลำดับ

ตารางที่ 5 ความน่าเชื่อถือในการสังเกตภายในสำหรับคะแนนความรุนแรงของภาพถ่ายรังสีทรวงอกครั้งที่ 1 และ 2 จากรังสีแพทย์ที่มีประสบการณ์ 11 ปี

โซน	สัมประสิทธิ์สหสัมพันธ์: ICC (95% CI)
รวม	0.97 (0.95-0.98)
A	0.95 (0.92-0.96)
B	0.94 (0.92-0.96)
C	0.95 (0.94-0.97)
D	0.92 (0.89-0.95)
E	0.92 (0.89-0.95)
F	0.89 (0.85-0.93)

การประเมิน intra-observer reliability โดยการแปลผลภาพถ่ายทางรังสีทรวงอก 2 ครั้งในช่วงเวลาต่างกันของ reader 1 พบว่าผล chest x-ray severity score ครั้งที่ 1 และ 2 ความน่าเชื่อถือมีความสอดคล้องกันในระดับดีมาก โดยรวมค่าความสอดคล้อง เท่ากับ 0.97 (ตารางที่ 5)



รูปที่ 3 Pearson correlation for chest x-ray severity score ครั้งที่ 1 และ 2 ของ reader 1

จากรูปที่ 3 พบว่า chest x-ray severity score ครั้งที่ 1 และ 2 ของ reader 1 มีความสัมพันธ์กันในระดับดีมาก ค่า Pearson correlation = 0.97, $p < 0.001$

วิจารณ์

การศึกษานี้พบว่าความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยโรคโควิด-19 โดยใช้ Brixia scoring system ของรังสีแพทย์ 3 ราย มีความสอดคล้องกันโดยรวมในระดับดีถึงดีมาก (weighted Cohen's $K = 0.79-0.90$) ความสอดคล้องในการแปลผลระหว่างรังสีแพทย์ที่มีประสบการณ์มาก ดีกว่ารังสีแพทย์ใหม่เล็กน้อย ปัจจัยที่อาจมีผลต่อการประเมินภาพถ่ายรังสีทรวงอกของรังสีแพทย์ใหม่คือความไม่คุ้นเคยกับลักษณะภาพถ่ายรังสีทรวงอกซึ่งส่วนใหญ่ถ่ายโดยเครื่องเอกซเรย์เคลื่อนที่ (portable X-ray machine) รวมถึงเทคนิคการปรับภาพของเจ้าหน้าที่รังสีเทคนิคของโรงพยาบาล

ส่วนความน่าเชื่อถือภายในผู้ประเมิน (intra-observer reliability) พบว่าความน่าเชื่อถือในการประเมินภาพถ่ายรังสีทรวงอกครั้งที่ 1 และ 2 ของ reader 1 มีความสอดคล้องกันในระดับดีมาก (ICC 0.97: 95%CI)

ค่าความน่าเชื่อถือภายในผู้ประเมิน (intra-rater reliability) ทั้ง 3 รายโดยรวมถือว่าสอดคล้องกันในระดับดี (ICC 0.88) เมื่อพิจารณาในแต่ละโซน พบว่าค่าความน่าเชื่อถืออยู่ในระดับดี ได้แก่ โซน B, C, D และ E (ICC 0.77-0.78: 95%CI)

ผลการศึกษานี้ใกล้เคียงกับการศึกษาของ Van Assen และคณะ¹³ ที่ศึกษาความสอดคล้องในการแปลผลภาพถ่ายทางรังสีระหว่างรังสีแพทย์ที่มีประสบการณ์และแพทย์ประจำบ้าน โดยพบว่าความสอดคล้องในการประเมินความรุนแรงของภาวะปอดอักเสบจากเชื้อโรคโควิด-19 ที่พบจากภาพถ่ายรังสีทรวงอก มีค่าความสอดคล้องกันในระดับดี ความสอดคล้องในการแปลผลระหว่างแพทย์ประจำบ้านต่ำกว่ารังสีแพทย์ที่มีประสบการณ์เล็กน้อย รวมถึงค่าความน่าเชื่อถือภายในผู้ประเมินก็อยู่ในเกณฑ์ดีเช่นกัน แต่ในการศึกษานี้

มีการกำหนดแนวทางการประเมินความผิดปกติของภาพถ่ายทางรังสี (severity classification) ขึ้นมาใช้เอง

การศึกษานี้สอดคล้องกับการศึกษาก่อนหน้านี้ของ Monaco และคณะ⁹ ที่ศึกษาความผิดปกติของปอดจากภาพถ่ายรังสีทรวงอกในผู้ป่วยโรคโควิด-19 โดยใช้ semiquantitative severity score และประเมินความสอดคล้องในการประเมินภาพถ่ายรังสีทรวงอกของรังสีแพทย์จาก 2 สถาบัน พบว่ามีความสอดคล้องกันในระดับดีถึงดีมากระหว่างรังสีแพทย์ในสถาบันเดียวกัน (up to $K = 0.872$ และ $K = 0.764$) ความสอดคล้องโดยรวมของรังสีแพทย์ทั้ง 2 สถาบันอยู่ระดับปานกลางถึงดีมาก (moderate ($K = 0.449$) to almost perfect ($K = 0.782$) inter-observer agreement) แต่การประเมินภาพถ่ายทางรังสีใช้ระบบการประเมินที่แตกต่างจากการศึกษานี้

ค่าความสอดคล้องในการศึกษานี้ดีกว่าการศึกษาก่อนหน้านี้ของ Cohen และคณะ¹⁴ ทำการศึกษาความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอก 94 ภาพ ระหว่างรังสีแพทย์ 2 ราย และแพทย์ประจำบ้าน 1 ราย พบว่าค่าความสอดคล้องอยู่ที่ 0.45 เมื่อประเมินโดยใช้ opacity score และ

ความสอดคล้องอยู่ที่ 0.71 เมื่อประเมินโดยใช้ extent score ผลการศึกษานี้ใกล้เคียงกับการศึกษาของ moncada และคณะ¹⁵ ซึ่งทำการศึกษา inter- and intra-observer variability ระหว่างรังสีแพทย์ในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยปอดอักเสบ (community acquired pneumonia) พบว่าความสอดคล้องระหว่างรังสีแพทย์ (inter-observer variability) เท่ากับ 0.72 และความน่าเชื่อถือภายในผู้ประเมิน (intra-observer variability) อยู่ระหว่าง 0.60-0.85 โดยใช้ semi-quantitative approach

สรุป

การศึกษานี้พบว่าความสอดคล้องในการแปลผลภาพถ่ายรังสีทรวงอกของผู้ป่วยโรคโควิด-19 โดยใช้ Brixia scoring system ของรังสีแพทย์ 3 ราย ทั้งรังสีแพทย์ที่มีประสบการณ์และรังสีแพทย์ใหม่ มีความสอดคล้องกันโดยรวมในระดับดีถึงดีมาก รวมถึงความน่าเชื่อถือภายในผู้ประเมิน (intra-observer reliability) พบว่าความน่าเชื่อถือมีความสอดคล้องกันในระดับดีมาก สามารถนำแนวทางการประเมินภาพถ่ายรังสีทรวงอกด้วย Brixia scoring system มาประยุกต์ใช้ในกลุ่มผู้ป่วยปอดอักเสบด้วยสาเหตุอื่นหรือนำมาใช้เมื่อเกิดโรคอุบัติใหม่ได้ในอนาคต รวมถึงสามารถนำข้อมูลมาพัฒนาต่อยอดสำหรับการสร้างระบบเทคโนโลยีปัญญาประดิษฐ์ (artificial intelligence: AI) ในการประเมินภาพถ่ายทางรังสีทรวงอก (chest x-ray) ของโรงพยาบาลต่อไป

กิตติกรรมประกาศ

ขอขอบคุณแพทย์หญิงบุปผารัตน์ อ่อนชมจันทร์และแพทย์หญิงภาวิ กุลพรหม ที่ให้ความร่วมมือในการเก็บข้อมูลการศึกษา เจ้าหน้าที่แผนกรังสีวิทยาและเจ้าหน้าที่ที่เกี่ยวข้องทุกท่านในการดูแลผู้ป่วยโรคโควิด-19 ของโรงพยาบาลชัยภูมิ

เอกสารอ้างอิง

1. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med* 2020;27(2):1-3. doi: 10.1093/jtm/taaa008
2. World Health Organization. Pneumonia of unknown cause – China. [cited Dec 13, 2020]. Available from: <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-causechina/en/>.
3. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, et al. Coronavirus disease 2019 (COVID-19): a perspective from China. *Radiology* 2020; 296(2):E15-E25. doi: 10.1148/radiol.202000490
4. Ng MY, Lee EYP, Yang J, Yang F, Li X, Wang H, et al. Imaging profile of the COVID-19 infection: radiologic findings and literature review. *Radiol Cardiothorac Imaging* 2020;2(1): e200034. doi: 10.1148/ryct.202000034
5. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology* 2020; 295(3):200463. doi: 10.1148/radiol.202000463
6. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest x-ray scoring system for quantifying and monitoring disease progression. *Radiol Med* 2020;125(5):509-13. doi: 10.1007/s11547-020-01200-3
7. World Health Organization. WHO global epidemiological surveillance standards for influenza. In. Edited by Influenza WHO WGESSf. Geneva, Switzerland: World Health Organization; 2013
8. Warren MA, Zhao Z, Koyama T, Bastarache JA, Shaver CM, Semler MW, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. *Thorax* 2018; 73(9):840-6. doi: 10.1136/thoraxjnl-2017-211280
9. Monaco CG, Zaottini F, Schiaffino S, Villa A, Della Pepa G, Carbonaro LA, et al. Chest x-ray severity score in COVID-19 patients on emergency department admission: a two-centre study. *Eur Radiol Exp* 2020;4(68):1-7. doi: 10.1186/s41747-020-00195-w
10. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17(1):101-10. doi: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.
12. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2): 155-63. doi: 10.1016/j.jcm.2016.02.012
13. Van Assen M, Zandehshahvar M, Maleki H, Kiarashi Y, Arleo T, Stillman AE, et al. COVID-19 pneumonia chest radiographic severity score: variability assessment among experienced and in-training radiologists and creation of a multireader composite score database for artificial intelligence algorithm development. *Br J Radiol* 2022;95 (1134):20211028. doi: 10.1259/bjr.20211028
14. Cohen JP, Dao L, Roth K, Morrison P, Bengio Y, Abbasi AF, et al. Predicting COVID-19 pneumonia severity on chest x-ray with deep learning. *Cureus* 2020;12(7): e9448. doi: 10.7759/cureus.9448
15. Moncada DC, Rueda ZV, Macías A, Suárez T, Ortega H, Vélez LA. Reading and interpretation of chest X-ray in adults with community-acquired pneumonia. *Braz J Infect Dis* 2011;15(6):540-6.

SMJ