

Inter- and intra-observer reliability of cobb angle measurement between orthopaedic residents and spine specialists

A. I. AL ALAWI¹, S. AL KALBANI¹, A. M. AL BARWANI¹, R. S. AL WESHAHI²

¹Department of Orthopaedic Surgery, Khoula Hospital, Directorate General of Khoula Hospital, Ministry of Health, Muscat, Oman;

²Oman Medical Specialty Board (OMSB), Muscat, Oman.

Correspondence at: Dr. Ahmed Mazin Al Barwani, Department of Orthopaedic Surgery, Khoula Hospital, Directorate, General of Khoula Hospital, Ministry of Health, Muscat, Oman. - E-mail: barwani4ever@gmail.com

ABSTRACT

Cobb angle measurement is essential for diagnosing and monitoring scoliosis; however, its accuracy varies with observer experience. In many healthcare systems, including Oman, initial measurements are commonly performed by residents and influence referral urgency, follow-up intervals, and bracing decisions, where even small inaccuracies may alter clinical pathways. This study evaluated the inter- and intra-observer reliability of Cobb angle measurement among orthopaedic residents and spine specialists. A prospective comparative reliability study was conducted including 18 OMSB orthopaedic residents (PGY2–5) and five spine specialists. Each participant measured Cobb angles on five standardised standing scoliosis radiographs on two occasions, four weeks apart. Inter- and intra-observer reliability were assessed using the intraclass correlation coefficient (ICC) with a two-way random-effects, single-measurement, absolute-agreement model [ICC(2,1)], and Bland–Altman plots were used to assess intra-observer reproducibility. Mean Cobb angles ranged from 45° to 56°. Inter-observer reliability was poor among residents (ICC = –0.219; 95% CI –1.41 to 0.485) and remained low among specialists (ICC = 0.120; 95% CI –2.28 to 0.898). Intra-observer variation averaged approximately 5°, with a Bland–Altman mean bias close to zero and limits of agreement of ±5%. Negative ICC values indicated that between-observer variability exceeded within-observer consistency. Residents demonstrated markedly poorer inter-observer reliability compared to specialists, highlighting the impact of clinical experience on accurate end-vertebra selection and angle measurement. Targeted training modules and structured exposure during residency may improve measurement consistency, referral accuracy, and overall scoliosis management pathways.

Keywords: Cobb angle, scoliosis, reliability, inter-observer, intra-observer, residency training.

INTRODUCTION

Scoliosis is a three-dimensional spinal deformity characterised by lateral curvature of more than 10° on standing radiographs, traditionally quantified using the Cobb angle, which remains the gold standard for assessment and follow-up. Despite its widespread use, Cobb angle measurement is subject to variation arising from end-vertebra selection, radiographic quality, and observer experience^{1,3}. Reported intra-observer variability typically ranges between 5° and 10°, with inter-observer variability often higher^{4–6}.

In clinical practice, spine specialists are responsible for definitive management decisions; however,

the initial Cobb angle estimation is almost always performed by residents or non-specialist clinicians. These measurements determine referral urgency, suitability for bracing, and intervals for surveillance imaging. Even modest inaccuracies during early triage may contribute to delayed identification of curve progression, unnecessary referrals, or inappropriate reassurance.

International studies demonstrate significant differences in measurement reliability between junior trainees and experienced spine surgeons^{7,8}. However, no published data are available from Oman or the Gulf region, despite the fact that Khoula Hospital receives the majority of scoliosis referrals nationwide.

Institution where the study was carried out: Khoula Hospital, Directorate General of Khoula Hospital, Ministry of Health, Muscat, Oman.

Given this gap, the present study aimed to evaluate the inter- and intra-observer reliability of Cobb angle measurement among orthopaedic residents and spine specialists.

MATERIALS AND METHODS

Study Design

A prospective comparative reliability study was conducted at Khoula Hospital, Muscat.

Ethical Approval

Approval was obtained from the Research and Ethics Committee, Directorate General of Khoula Hospital, Ministry of Health, Oman (Approval No. MOH/DGKH/REC/22/10/25831; 13 October 2022). All radiographs were anonymised.

Participants

Eighteen OMSB orthopaedic residents (PGY2–5) and five spine specialists (2–20 years' experience) participated. First-year residents were excluded due to limited exposure to spinal pathology.

Radiographs and Measurement Procedure

Five standing AP scoliosis radiographs of varying curve magnitudes were selected. A standardised instructional session on Cobb measurement was provided to all available participants; specialists who could not attend received a brief refresher immediately before measurement.

Radiographs were reviewed on identical PACS workstations. Each participant independently measured all five radiographs. Approximately four weeks later, participants repeated the measurements in the same order to assess intra-observer reliability.

The number of radiographs was limited to five to minimise recall bias and because reliability studies typically achieve stable ICC values with 4–6 images. This design aligns with similar published reliability studies.

Sample Size Justification

A minimum of 15–20 raters is generally adequate to estimate ICC with acceptable precision in reliability studies. With 23 participants, this sample size provides stable ICC estimates and complies with methodological recommendations for reliability research.

Statistical Analysis

For each observer, mean Cobb angle and standard deviation (SD) were calculated. Inter- and intra-

observer reliability were assessed using the intraclass correlation coefficient (ICC) with a two-way random-effects, single-measurement, absolute-agreement model [ICC(2,1)], suitable when both raters and radiographs are considered random samples.

ICC thresholds followed Koo and Li (2016):

- Poor: <0.50
- Moderate: 0.50–0.75
- Good: 0.75–0.90
- Excellent: >0.90

Bland–Altman plots were constructed to evaluate intra-observer agreement. Bias (mean difference) and limits of agreement (mean $\pm$ 1.96 SD) were reported. Statistical analyses were conducted using SPSS Statistics v27 (IBM Corp., Armonk, NY, USA).

RESULTS

All 23 participants completed both measurement sessions. Mean Cobb angles ranged from 45° to 56° across raters (Figure 1), with residents showing wider variability than specialists (Table I). Residents showed large variability, with SD values ranging from 3° to 16°, indicating wide differences in angle selection and measurement. Specialists demonstrated narrower SD ranges (5°–8°), reflecting closer agreement.

Inter-observer reliability among residents was poor (ICC = –0.219; 95% CI –1.41 to 0.485), indicating substantial inconsistency. Specialists demonstrated slightly higher but still poor reliability (ICC = 0.120; 95% CI –2.28 to 0.898). Negative ICC values suggest that variability between observers was greater than variability within individual observers—a recognised statistical phenomenon when agreement is extremely low.

Intra-observer reliability was more consistent across both groups, with variation of approximately 5°. Bland–Altman analyses showed a mean bias close to zero and limits of agreement of $\pm$ 5°, consistent with published reproducibility values for Cobb measurement. No clear association was found between curve magnitude and measurement variability (Figure 2 and Figure 3).

DISCUSSION

This study demonstrated poor inter-observer reliability among orthopaedic residents compared with spine specialists, confirming the strong effect of observer experience on accurate identification of end

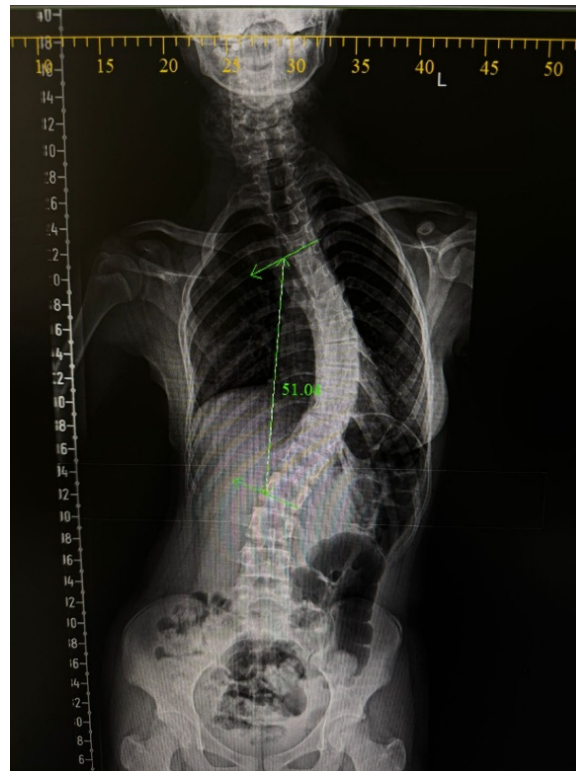


Fig. 1 — Standing AP radiograph demonstrating Cobb angle measurement.

Table I. — Average Cobb Angle, Standard Deviation (SD), and ICC.

Examiner	Mean Cobb Angle (°)	SD (°)	ICC	95% CI
Residents (1–18)	46.4–56.4	3–16.8	–0.219	–1.41 to 0.485
Specialists (19–23)	50.4–53.8	5.6–7.7	0.120	–2.28 to 0.898

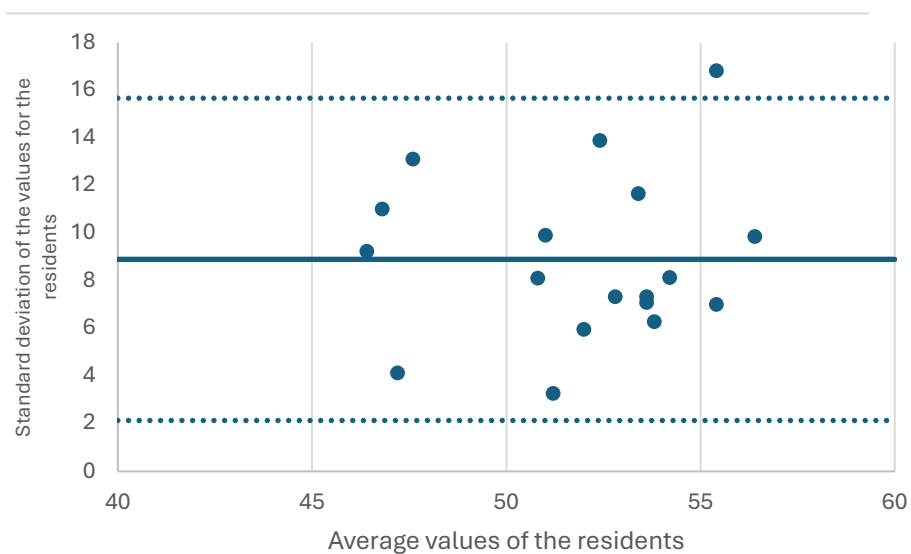


Fig. 2 — Scatter plot showing residents' mean Cobb angles vs. SD.

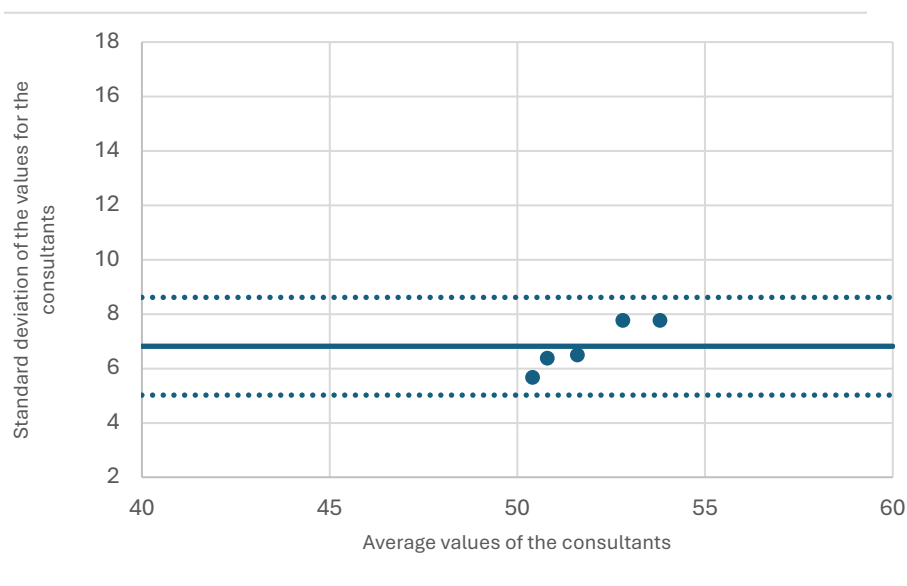


Fig. 3 — Scatter plot showing specialists' mean Cobb angles vs. SD.

vertebrae and angle construction. These findings align with prior research reporting lower reliability among less experienced examiners^{3,7,8}.

Although digital PACS tools enhance precision, recent studies (2020–2023) show that they do not eliminate observer-dependent variability^{9–11}. Our results support this observation: intra-observer reliability was acceptable across both groups, but inter-observer variability—especially among residents—remained markedly high.

The negative ICC values recorded among residents indicate disagreement exceeding what would be expected by chance. Such values can occur when between-rater variability surpasses within-rater consistency, and should be interpreted not as “negative reliability,” but as an indication of substantial disagreement.

The clinical implications are noteworthy. Khoula Hospital functions as a national referral centre, and inaccurate Cobb angle assessment at primary or secondary levels can lead to delayed progression detection, inappropriate bracing decisions, or unnecessary referrals. Structured scoliosis modules within residency training, including supervised end-vertebra identification and hands-on digital measurement practice, may significantly improve reliability.

This study has several limitations. Only five radiographs were evaluated, which, although consistent with common reliability study designs, may limit the representation of the full spectrum of scoliosis curve patterns. Variability in attendance

during the instructional session may also have introduced differences in baseline understanding and influenced measurement performance. Furthermore, the relatively small number of spine specialists included may have limited the precision and generalisability of the specialist ICC estimates.

CONCLUSION

Cobb angle measurement reliability is substantially lower among residents than specialists, with significant inter-observer disagreement. Structured training and supervised measurement practice may improve reproducibility and enhance referral accuracy within national scoliosis care pathways.

Funding: No external funding was received.

Conflict of Interest: The authors declare no conflicts of interest.

REFERENCES

1. Youssef JA, Orndorff DO, Patty CA, Scott MA, Price HL, Hamlin LF, et al. Current status of adult spinal deformity. *Global Spine J.* 2013;3:51–62.
2. Ritter R, Nagasse Y, Ribeiro I, Yamazato C, de Oliveira FM, Kusabara R. Comparison of Cobb angle measurement in scoliosis by residents and spine experts. *Coluna/Columna.* 2016;15(1):13–16.
3. Mehta SS, Modi HN, Srinivasalu S, Chen T, Suh SW, Yang JH, et al. Interobserver and intraobserver reliability of Cobb angle measurement: endplate versus pedicle as landmarks. *J Pediatr Orthop.* 2009;29(7):749–754.
4. Stokes IA, Aronson DD, Ronchetti PJ, Labelle H, Dansereau J. Re-examination of Cobb and Ferguson angles: bigger is not always better. *J Spinal Disord.* 1993;6(4):333–338.

5. Dimar JR, Carreon LY, Labelle H, Djurasovic M, Weidenbaum M. Intra- and inter-observer reliability of determining radiographic sagittal parameters of the spine and pelvis using a manual and a computer-assisted methods. *Eur Spine J*. 2008;17:1373–1379.
6. Wang J, Zhang J, Xu R, Chen TG, Zhou KS, Zhang HH. Measurement of scoliosis Cobb angle by end vertebra tilt method. *J Orthop Surg Res*. 2018;13:223.
7. Choudhry MN, Ahmad Z, Verma R. Adolescent idiopathic scoliosis. *Open Orthop J*. 2015;10:143–154.
8. Greiner KA. Adolescent idiopathic scoliosis: radiologic decision-making. *Am Fam Physician*. 2002;65(9):1817–1822.
9. Lucasti C, Haider MN, Marshall IP, Thomas R, Scott MM, Ferrick MR. Inter- and intra-reliability of Cobb angle measurement in pediatric scoliosis using PACS (picture archiving and communication systems methods) for clinicians with various levels of experience. *Ann Spinal Joint*. 2023 Feb 13;7:1.
10. Hurtado-Avilés J, et al. Validity and reliability of Cobb angle measurements using digital radiographic images: a comparative study between experienced and novice observers. *Int J Environ Res Public Health*. 2022;19(8):4655.
11. Prestigiacomo FG, et al. Intra- and interobserver reliability of Cobb angle measurements in adolescent idiopathic scoliosis. *Spine Deform*. 2021;9(5):1227–1233.