

Reliability of cervical lordosis measurement techniques on long-cassette radiographs

Piotr Janusz^{1,2} · Marcin Tyrakowski³ · Hailong Yu⁴ · Kris Siemionow¹

Received: 26 June 2015 / Revised: 26 November 2015 / Accepted: 26 November 2015 / Published online: 26 December 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Purpose Lateral radiographs are commonly used to assess cervical sagittal alignment. Three assessment methods have been described and are commonly utilized in clinical practice. These methods are described for perfect lateral cervical radiographs, however in everyday practice radiograph quality varies. The aim of this study was to compare the reliability and reproducibility of 3 cervical lordosis (CL) measurement methods.

Methods Forty-four standing lateral radiographs were randomly chosen from a lateral long-cassette radiograph database. Measurements of CL were performed with: Cobb method C2–C7 (CM), C2–C7 posterior tangent method (PTM), sum of posterior tangent method for each segment (SPTM). Three independent orthopaedic surgeons measured CL using the three methods on 44 lateral radiographs. One researcher used the three methods to measure CL three times at 4-week time intervals. Agreement between the methods as well as their intra- and interobserver reliability were tested and quantified by intraclass correlation

coefficient (ICC) and median error for a single measurement (SEM). ICC of 0.75 or more reflected an excellent agreement/reliability. The results were compared with repeated ANOVA test, with $p < 0.05$ considered as significant.

Results All methods revealed excellent intra- and interobserver reliability. Agreement (ICC, SEM) between three methods was (0.89°, 3.44°), between CM and SPTM was (0.82°, 4.42°), between CM and PTM was (0.80°, 4.80°) and between PTM and SPTM was (0.99°, 1.10°). Mean values CL for a CM, PTM, SPTM were $10.5^\circ \pm 13.9^\circ$, $17.5^\circ \pm 15.6^\circ$ and $17.7^\circ \pm 15.9^\circ$ ($p < 0.0001$), respectively. The significant difference was between CM vs PTM ($p < 0.0001$) and CM vs SPTM ($p < 0.0001$), but not between PTM vs SPTM ($p > 0.05$).

Conclusions All three methods appeared to be highly reliable. Although, high agreement between all measurement methods was shown, we do not recommend using Cobb measurement method interchangeably with PTM or SPTM within a single study as this could lead to error, whereas, such a comparison between tangent methods can be considered.

✉ Piotr Janusz
mdpjanusz@gmail.com

¹ Department of Orthopaedic Surgery, University of Illinois at Chicago, 835 South Wolcott Ave, Room E-270, Chicago, IL 60612, USA

² Department of Spine Disorders and Pediatric Orthopedics, University of Medical Sciences, Poznan, Poland

³ Department of Orthopaedics, Pediatric Orthopaedics and Traumatology, The Centre of Postgraduate Medical Education in Warsaw, Konarskiego 13, 05-400 Otwock, Poland

⁴ Department of Orthopedics, General Hospital of Shenyang Military Area Command of Chinese PLA, Shenyang 110016, People's Republic of China

Keywords Cervical lordosis measurement · Cobb method · Cervical lordosis · Harrison method

Introduction

Spinal balance is critical for physiologic function and low energy expenditure [1]. Sagittal cervical alignment is one of the most important parameters in management of cervical spine disorders [2]. It is thus crucial to use a reliable and reproducible measurement method; one that allows to properly assess the course of the disease and the results of treatment [2]. Cervical lordosis (CL) is the most commonly

used cervical parameter by surgeons and researchers [3]. Three distinct cervical lordosis assessment methods have been described [4]. The Cobb method using lines perpendicular to C2 and C7 vertebral distal end plate lines was primarily described to evaluate scoliotic curves [4–6]. In some cases it is also considered as Cobb angle between C1 and C7 vertebrae [4]. The Harrison posterior tangent method calculates sum of segmental angles measured between lines parallel to the posterior surface of each cervical vertebral bodies from C2 to C7 for an overall cervical curvature angle [4, 7]. In the Jackson method the angle between lines parallel to the posterior surface of the C7 and C2 vertebral bodies is measured [4, 8]. This method was also used by Gore et al. and is often known as a Gore method [9].

Although Harrison et al. suggested that the Harrison method may provide the best measurement of CL [7], the Cobb method is still the most widely used [4]. Reliability of these three methods was evaluated using lateral cervical radiographs and an attempt to establish the best method of measurement was previously made [2, 3, 7]. However, whole-spine lateral radiographs are a key image used in evaluation of global spinal sagittal alignment [10]. According to Park et al. there are possible difference in radiological parameter measurements between lateral cervical radiographs and whole-spine lateral radiograph [10].

The aim of this study was to evaluate agreement between the three methods of CL measurements, as well as their reliability and reproducibility on standing long-cassette lateral radiographs of the spine.

Methods

After obtaining Institutional Review Board approval, a database of standard standing digital long-cassette lateral radiographs of the whole spine taken between June 2009

and June 2014 was retrospectively reviewed. Forty-four standing lateral radiographs were randomly chosen from the radiograph database.

Similar radiologic protocol was used during the entire study period. Lateral radiographs were obtained with each subject standing in natural position, with horizontal gaze (patients looked at the point on the wall at the sight level 2 m in front of them), shoulders flexed 30°–45° and the elbows slightly flexed with the hands resting on a support. Hips and knees were in full extension. The radiographs covered the pelvis with the hips, the whole spine and the cranium to the level of the external auditory meatus and the lower margin of the orbit.

On each of the lateral radiographs the angle of CL was measured using the following methods:

1. Cobb method C2–C7 (CM)—the angle between the inferior end plate of C2 and the inferior end plate of C7 [4–6] (Fig. 1a).
2. C2–C7 posterior tangent method (PTM)—described as the Jackson method or the Gore angle [4, 8, 9]—the angle between the line sustained by posterior margin of C2 vertebral body and the posterior margin of C7 vertebral body (Fig. 1b).
3. Sum of posterior tangents method (SPTM) for segments C2–C3, C3–C4, C5–C6, C6–C7—described as the Harrison method [4, 7]—sum of the angles measured with the sagittal tangent method at five levels: C2–C3, C3–C4, C4–C5, C5–C6, C6–C7 (Fig. 1c).

Lordotic CL angles were presented as positive values, and kyphotic CL angles were presented as negative values.

All of the radiographs were downloaded from the Centricity PACS system (General Electric Medical Systems, Centricity PACS Radiology RA1000 Workstation; General Electric Healthcare, Barrington, IL) as bitmap images and analyzed quantitatively with Surgimap Spine Software (Surgimap, New York, USA).

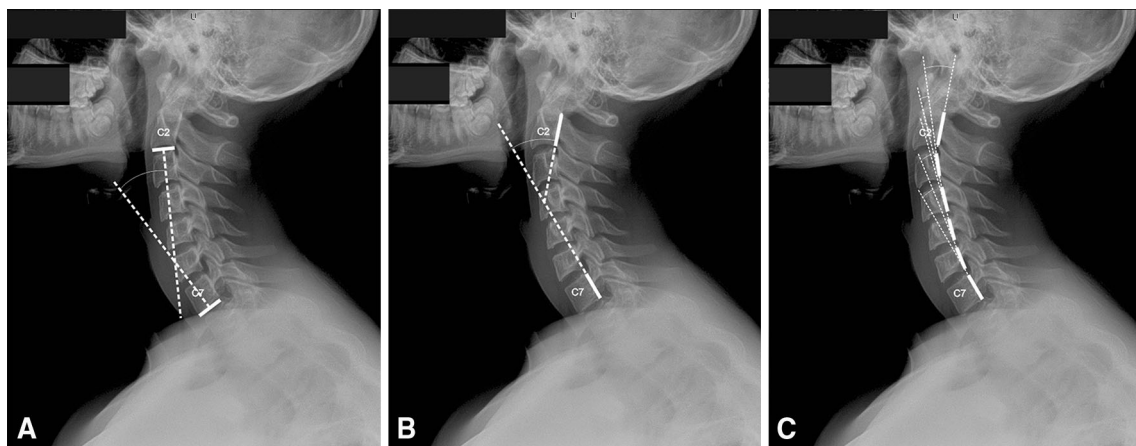


Fig. 1 Three methods of cervical lordosis measurements: **a** Cobb method C2–C7 (CM); **b** C2–C7 posterior tangent method (PTM); **c** sum of posterior tangents method (SPTM) for segments C2–C3, C3–C4, C5–C6, C6–C7

Evaluation of the intraobserver reproducibility of the three methods of CL measurements

The measurements were performed on 44 radiographs by one researcher (orthopedic spine surgeon with 5 years of experience) 3 times at 4-week intervals. The order of the radiographs in the second and third series of measurements was different and random. The intraobserver reproducibility was tested and quantified by intraclass correlation coefficient (ICC) and median error for a single measurement (SEM) [11].

Evaluation of the interobserver reliability of three methods of CL measurements

Three independent researchers (orthopedic spine surgeons with 10, 6 and 5 years of experience) measured CL on the same 44 radiographs once with each of three methods tested. The interobserver reliability was tested and quantified by intraclass correlation coefficient (ICC) and median error for a single measurement (SEM) [11].

Evaluation of agreement between the three methods of CL measurements

The evaluation of agreement between the three methods of CL assessment was based on the measurements performed by one randomly chosen researcher (orthopaedic spine surgeon with 6 years of experience) on 44 radiographs.

Agreement between the methods was quantified by the intraclass correlation coefficient (ICC) and the median error for a single measurement (SEM) [11].

Statistical analysis

The data were analyzed using the JMP 10.0.2 (SAS Institute Inc, Cary, NC) statistical software and in Microsoft Office Excel 2007 (Microsoft, Redmond, WA). The ICC value of less than 0.40 indicated poor agreement, 0.40–0.75 indicated fair to good agreement, and values greater than 0.75 reflected excellent agreement [12]. To estimate the sample size needed to test the agreement between the three methods evaluated, as well as the intraobserver reproducibility and interobserver reliability for all of the methods we treated the ICC value greater than 0.7 (with its 95 % confidence interval of 0.55–0.85) as having an acceptable reproducibility for a research tool [13]. The minimum number of subjects to test the agreement, intraobserver reproducibility and interobserver reliability in our setting was 44 [14]. Randomizations were performed by use of RAND function in Microsoft Office Excel 2007.

For each parameter the mean values, standard deviation, and the values range were established. Normal distribution of data was analyzed with the Shapiro–Wilk test. The results were compared with repeated ANOVA test with Bonferroni correction, with $p < 0.05$ considered as significant.

Results

Among the evaluated patients there were 13 males and 31 females, with a mean age of 15.8 ± 3.7 years.

Evaluation of the intraobserver reproducibility of the three methods of CL measurements

Intraobserver reliability was excellent for all of the methods tested with ICC = 0.96 and SEM = 2.06° for CM, ICC = 0.96 and SEM = 1.99° for PTM, and ICC = 0.96 and SEM = 1.98° for SPTM, Table 1.

Evaluation of the interobserver reliability of the three methods of CL measurements

Interobserver reliability was excellent for all of the methods tested with ICC = 0.92 and SEM = 2.71° for CM, ICC = 0.94 and SEM = 2.62° for PTM, and ICC = 0.93 and SEM = 2.78° for SPTM, Table 1.

The intraobserver reliability of segmental CL measured with SPTM was excellent at all levels from C2 to C7, with the lowest value at C5–C4 level, Table 2. The interobserver reliability of segmental CL measured with SPTM was excellent at all levels from C2 to C7, with the lowest value at C5–C4 level, Table 2.

Evaluation of agreement between the three methods of CL measurements

The overall agreement between three methods tested was excellent with ICC = 0.89 and SEM = 3.44° . In pairs

Table 1 Reliability of segmental cervical lordosis measurements according to Harrison method

Method	Intraobserver reproducibility		Interobserver reliability	
	ICC	SEM ($^\circ$)	ICC	SEM ($^\circ$)
CM	0.96	2.06	0.92	2.71
PTM	0.96	1.99	0.94	2.62
SPTM	0.96	1.98	0.93	2.78

ICC Intraclass correlation coefficient, SEM median error for a single measurement, CM Cobb method, PTM C2–C7 posterior tangent method, SPTM sum of posterior tangents for segments C2–C3, C3–C4, C5–C6, C6–C7

Table 2 Reliability of segmental cervical lordosis measurements according to Harrison method

Segment	Intraobserver		Interobserver	
	ICC	SEM (°)	ICC	SEM (°)
C7–C6	0.88	1.55	0.91	1.52
C6–C5	0.89	1.48	0.92	1.66
C5–C4	0.83	1.77	0.77	1.87
C4–C3	0.88	1.79	0.83	1.96
C3–C2	0.86	1.54	0.80	2.09

ICC Intraclass correlation coefficient; SEM median error for a single measurement

Table 3 Agreement between three methods of cervical lordosis measurements

Methods	ICC	SEM (°)
Overall agreement (CM-PTM-SPTM)	0.89	3.44
CM-SPTM	0.82	4.42
CM-PTM	0.80	4.80
PTM-SPTM	0.99	1.10

ICC Intraclass correlation coefficient, SEM median error for a single measurement, CM Cobb method, PTM C2–C7 posterior tangent method, SPTM sum of posterior tangents for segments C2–C3, C3–C4, C5–C6, C6–C7

comparison revealed excellent agreement for all of the methods with $ICC \geq 0.80$ and $SEM \leq 4.80^\circ$, Table 3.

Mean values CL for a Cobb method, tangent method, tangent sum method were $10.5^\circ \pm 13.9^\circ$, $17.5^\circ \pm 15.6^\circ$ and $17.7^\circ \pm 15.9^\circ$, respectively. The values of CL of each patients measured with three methods are presented in Fig. 2.

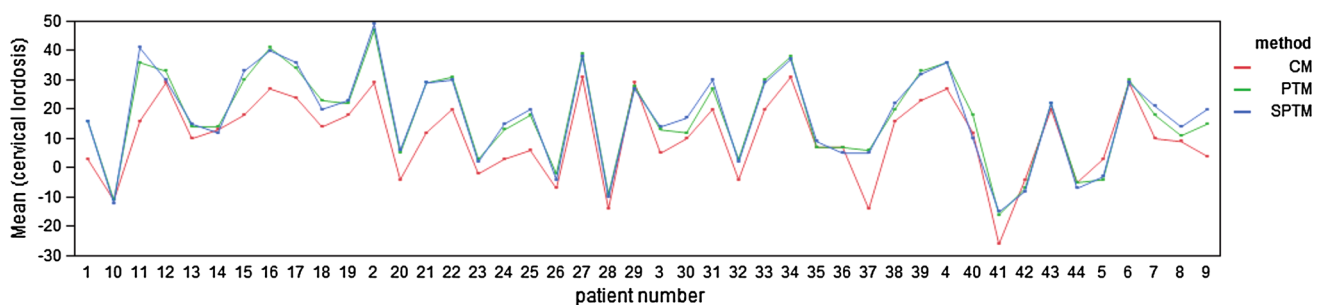
The difference between three methods was statistically significant ($p < 0.0001$, $F 48.43$). The pair comparison with Bonferroni correction revealed significant difference was between Cobb method versus tangent method ($p < 0.0001$) and Cobb method versus tangent sum method ($p < 0.0001$), but not between tangent method versus tangent sum method ($p > 0.05$).

Discussion

We present a comparison of three methods of CL measurements on standing long-cassette radiographs of the spine. Such an analysis has never been previously published.

Long-cassette lateral radiographs are important in global sagittal balance assessment [10]. On such radiographs cervical alignment can be measured and the relationship between CL and other spine segments can be established. Evaluating CL on long-cassette lateral radiographs may avoid additional radiation exposition associated with obtaining dedicated cervical radiographs. This is important in every patient, however, especially in children and the adolescent population [15].

Discrepancies between lateral cervical radiographs and long-cassette whole-spine radiographs in spinal parameters were reported. Park et al. described significant difference between CL values on lateral cervical radiographs and long-cassette whole-spine radiographs in the same individuals [10]. Body positions, arm placement, and focus distance are usually different between plain cervical and long-cassette, whole-spine radiographs [10, 16]. Taking into consideration the previously mentioned reports and the fact that previous studies concerning the reliability of CL measurement methods were based on lateral cervical radiographs, the evaluation of the measurements reliability for long-cassette whole-spine radiographs was needed. The three evaluated methods of measuring CL, namely the Cobb (CM), the Jackson (PTM) and the Harrison (SPTM) method proved to be reliable, which is in line with data presented for lateral cervical radiographs [3, 7]. In neither the intraobserver agreement nor in interobserver agreement evaluation have we found a predominance of any of the evaluated methods. In segmental analysis, all measurements in intra and interobserver evaluation showed excellent agreement with slightly lower ICC values at C5–C4, C4–C3 and C3–C2 levels. This partially stays in line with Harrison's et al. paper who reported lower reliability for C3–C2, C4–C3 and C7–C6 levels [7].

**Fig. 2** The values of CL of each patients measured with three methods are presented

We initially expected, that the SPTM method could have a slightly lower ICC due to the number of calculated segments (separate measures), which were summed and with a possibility of error at each level. Despite this, the results were within an excellent agreement interval.

Park et al. performed interclass correlation coefficient calculation for Cobb and Gore (PTM) measurements method performed by two researchers on cervical lateral radiographs and whole-spine lateral radiographs with both demonstrating excellent ICC [10]. In interobserver evaluation, Cote et al. reported that the Cobb method had an ICC of 0.96 [17]. There are two other studies describing CL measurements on plain lateral radiographs performed by Ohara et al. [3] and Silbert et al. [2]. In both studies, the conclusions are in line with ours, however, due to differences in statistical method used to evaluate their results (Pearson correlation coefficient), a direct comparison of results is not possible.

When evaluating studies focusing on clinical results and not the measurement methodology, the ICC may be lower, than presented in studies describing measuring methods [7, 10]. Park et al. performed cervical lordosis measurement according to Cobb method in different age groups on full length spine radiographs [18]. The ICC for the Cobb angle was 0.777 for the intraobserver reliability and 0.672 for the interobserver reliability [18].

The complexity of the shape of the cervical vertebrae, curvature of surface of the vertebral end plates and presence of uncinata processes can be confusing at radiographs when 3D structure is presented as a two dimensional picture and all the structures overlap each other [19, 20]. In this situation, radiological image of the posterior surface of the vertebrae seems to be more clear and less affected by overlapping structures. Thus we expected that the Cobb method may be less replicable than the posterior tangents methods. However, this has not been reflected by the results of our study. What is more, contrary to studies concerning cervical radiographs, ICC for the Cobb method and for other methods were similar, when in other studies ICC for Cobb method was slightly lower than for Jackson method [7, 10].

Currently, there is not a standard cervical alignment assessment method. Each method has proponents and opponents and all methods can be found in published data [2, 3, 7, 10]. It is important to know if the results can be compared in a reliable manner or used interchangeably without significant bias. Thus we performed agreement calculation between evaluated methods.

When evaluating results, it is important to not only focus on the analysis of the ICC value but also graph analysis and SEM evaluation. Considering only ICC value can result in an improper conclusion being reached that all methods are in excellent agreement and can be used interchangeably.

However, when we evaluate SEM perception of these result can be different. Since agreement between SPTM and PTM SEM is low (1.10°), we could incorrectly assume that this value would not have an important clinical effect. However, in agreement analysis between the Cobb methods and both tangent methods, SEM is four times larger. Such a high value of SEM, especially in relation to mean CL suggests that we should not consider using Cobb method interchangeably with both tangent methods, because it could be a source of substantial error.

Harrison et al. suggested that Cobb method underestimate CL [7]. In published data CL calculated with Cobb angle is lower than in posterior tangents methods in the same patients [3, 7, 10]. In our study the mean Cobb angle was approximately 7° lower than in posterior tangent methods, when the difference between both tangents methods was 0.2° . Thus we wanted to assess if CL values achieved with different methods differ significantly. Our concerns were confirmed in repeated ANOVA calculations.

One of the limitations of this study could be that we used radiographs without dividing them into subgroups according to age or disease. However, the authors' idea in this study was to assess random radiograms typically used in everyday practice, regardless to listed factors. Further studies are needed within specialized subgroups. Another limitation is the fact, that the ideal method of measurement requires very distinct vertebral contours, however, in lateral long-cassette radiograms this is not always the case. What is more, often in lateral long-cassette radiograms the vertebral borders are more blurred than in lateral cervical radiographs. Actually, the idea of this study was based on possible difference between these two types of radiograms and the possible consequence in measurement results.

To our knowledge this is the first study comparing CL measurements methods on standing long-cassette whole-spine radiograms. The strong side of this study is the method of analysis based on interclass correlation coefficient, median error for single measurement and Bland and Altman idea of comparison between measuring methods [21, 22]. The fact that measurements were performed with widely used free software (Surgimap Spine) can be an additional advantage for surgeons and researchers. Analysis of published results without taking into consideration the measurement method might lead not only to scientific bias but also therapeutic miscalculation.

Conclusions

All three methods appeared to be highly reliable. Although, high agreement between all measurement methods was shown, we do not recommend using Cobb measurement method interchangeably with PTM or SPTM within a

single study as this could lead to error, whereas, such a comparison between tangent methods can be considered.

Compliance with ethical standards

Conflict of interest None.

References

- Abelin-Genevois K, Idjerouidene A, Roussouly P, Vital JM, Garin C (2014) Cervical spine alignment in the pediatric population: a radiographic normative study of 150 asymptomatic patients. *Eur Spine J* 23(7):1442–1448. doi:10.1007/s00586-013-3150-5
- Silber JS, Lipetz JS, Hayes VM, Lonner BS (2004) Measurement variability in the assessment of sagittal alignment of the cervical spine: a comparison of the gore and cobb methods. *J Spinal Disord Tech* 17(4):301–305
- Ohara A, Miyamoto K, Naganawa T, Matsumoto K, Shimizu K (2006) Reliabilities of and correlations among five standard methods of assessing the sagittal alignment of the cervical spine. *Spine* 31(22):2585–2591. doi:10.1097/01.brs.0000240656.79060.18 (discussion 2592)
- Scheer JK, Tang JA, Smith JS, Acosta FL Jr, Protosaltis TS, Blondel B, Bess S, Shaffrey CI, Deviren V, Lafage V, Schwab F, Ames CP, International Spine Study Group (2013) Cervical spine alignment, sagittal deformity, and clinical implications: a review. *J Neurosurg Spine* 19(2):141–159. doi:10.3171/2013.4.SPINE.12838
- Cobb J (1948) Outline for the study of scoliosis. *Am Acad Orthop Surg Instr Course Lect* 5:261–275
- Sevastikoglou JA, Bergquist E (1969) Evaluation of the reliability of radiological methods for registration of scoliosis. *Acta Orthop Scand* 40(5):608–613
- Harrison DE, Harrison DD, Cailliet R, Troyanovich SJ, Janik TJ, Holland B (2000) Cobb method or Harrison posterior tangent method: which to choose for lateral cervical radiographic analysis. *Spine* 25(16):2072–2078
- J R (1958) *The cervical syndrome*, 2nd edn. Charles C Thomas, Springfield
- Gore DR, Sepic SB, Gardner GM (1986) Roentgenographic findings of the cervical spine in asymptomatic people. *Spine* 11(6):521–524
- Park SM, Song KS, Park SH, Kang H, Daniel Riew K (2015) Does whole-spine lateral radiograph with clavicle positioning reflect the correct cervical sagittal alignment? *Eur Spine J* 24(1):57–62. doi:10.1007/s00586-014-3525-2
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–428
- Streiner DL, Norman GR (2008) *Health measurement scales: a practical guide to their development and use*, 4th edn. Oxford University Press, Oxford
- Keszei AP, Novak M, Streiner DL (2010) Introduction to health measurement scales. *J Psychosom Res* 68(4):319–323. doi:10.1016/j.jpsychores.2010.01.006
- Zou GY (2012) Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med* 31(29):3972–3981. doi:10.1002/sim.5466
- Presciutti SM, Karukanda T, Lee M (2014) Management decisions for adolescent idiopathic scoliosis significantly affect patient radiation exposure. *Spine J* 14(9):1984–1990. doi:10.1016/j.spinee.2013.11.055
- Horton WC, Brown CW, Bridwell KH, Glassman SD, Suk SI, Cha CW (2005) Is there an optimal patient stance for obtaining a lateral 36° radiograph? A critical comparison of three techniques. *Spine* 30(4):427–433
- Cote P, Cassidy JD, Yong-Hing K, Sibley J, Loewy J (1997) Apophysial joint degeneration, disc degeneration, and sagittal curve of the cervical spine. Can they be measured reliably on radiographs? *Spine* 22(8):859–864
- Park MS, Moon SH, Lee HM, Kim SW, Kim TH, Lee SY, Riew KD (2013) The effect of age on cervical sagittal alignment: normative data on 100 asymptomatic subjects. *Spine* 38(8):E458–E463. doi:10.1097/BRS.0b013e31828802c2
- Tanaka N, Fujimoto Y, An HS, Ikuta Y, Yasuda M (2000) The anatomic relation among the nerve roots, intervertebral foramina, and intervertebral discs of the cervical spine. *Spine* 25(3):286–291
- Bland JH, Boushey DR (1990) Anatomy and physiology of the cervical spine. *Semin Arthritis Rheum* 20(1):1–20
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476):307–310
- Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19(1):231–240. doi:10.1519/15184.1