Interobserver reliability and intraobserver reproducibility of powers ratio for assessment of atlanto-occipital junction: comparison of plain radiography and computed tomography

Gang Li · Peter Passias · Michal Kozanek · Brian D. Shannon · Guoan Li ·
Fernando Villamil · Christopher M. Bono · Mitchel Harris · Kirkham B. Wood

Received: 15 July 2008 / Revised: 15 December 2008 / Accepted: 27 December 2008 / Published online: 23 January 2009
© Springer-Verlag 2009

Abstract Powers ratio, as assessed on plain radiographs or computed tomography (CT) images, appears to have clinical and prognostic value. To date, the validation of this assessment tool has been limited to a small number of observers at a single site. No study has examined the intraobserver reproducibility and interobserver reliability of the Powers ratio measurement on plain radiographs or CT images among a large cohort of spine surgeons. This type of validation is critical to allow for the broader use of the Powers ratio methodology in research studies and clinical applications. Plain radiographs and spiral CT images of the cervical spine of 32 patients were assessed, and the Powers ratio was determined by five spine surgeons. Each surgeon performed three readings, 7 months apart. In the first round of measurements, the observers used only the Powers' method of instruction. The second and third measurement sets were obtained after an interactive teaching session on the methodology. The order of the images was altered for

the second and third set of measurements. The coefficient of variation (Cv) was calculated to determine the intraobserver repeatability and interobserver reliability for each imaging technique. A Bland-Altman plot was then used to assess the agreement between the two imaging techniques. For interobserver reliability, the mean Cv of the Powers ratio was 9.09 and 4.31% for plain radiographs and CT, respectively. The Cv mean value for intraobserver reproducibility averaged 4.95% (range 1.39–9.08) when CT scans were used and 14.17% (range 7.54–34.30) when plain radiographs were used. For intraobserver reproducibility, the lowest and highest Cv mean value of five raters was 1.39 and 9.08% using CT scans and 7.54 and 34.3% using plain radiographs. The Bland-Altman plot, demonstrated that the two methods were in close agreement on the -0.8 and 0.89% interval for limits of agreement ($\text{bias} \pm 1.96\sigma$). The intraobserver reproducibility and interobserver reliability of Powers ratio measurement was acceptable ($<5\%$) with CT scans but not with plain radiographs. However, despite the statistically inferior reliability and repeatability, the Bland-Altman plot analysis showed that given the -0.8 and 0.89% limits of agreement, the two methods may be used interchangeably in clinical practice.

G. Li · P. Passias · M. Kozanek · B. D. Shannon · G. Li ·
F. Villamil · K. B. Wood (✉)
Department of Orthopedics, Spine Service,
Massachusetts General Hospital, Boston,
MA 02114, USA
e-mail: kbwood@partners.org

G. Li
Orthopaedic Department,
Peking University Third Hospital,
100083 Beijing, China

C. M. Bono · M. Harris
Department of Orthopedics, Spine Service,
Brigham and Women's Hospital,
Boston, MA 02115, USA

Keywords Powers ratio · Interobserver reliability ·
Intraobserver reproducibility · Atlanto-occipital junction

Introduction

Measurements obtained from both plain radiographs and computed tomography (CT) scans have been used to assess upper cervical spine injuries with the use of the Powers ratio (a ratio of the length of a line extending from the

basion to the posterior arch of the atlas divided by that of a line extending from the opisthion to the anterior arch of the atlas (Fig. 1) [9]. This is particularly useful for the diagnosis of atlanto-occipital dislocation (AOD) [2, 5–8, 11]. A strong correlation between the AOD and Powers ratio has been supported by both experimental and clinical studies using various methods including X-ray, CT and three-dimensional CT [2, 7, 11].

Most previous clinical studies have focused on the qualitative comparison of the Powers ratio [9], and other measurement techniques e.g. X-line method, condylar gap and Harris method [2, 6, 7]. However, very few of them have focused on intraobserver reproducibility and interobserver reliability. This study was undertaken via a multicentric team of fellowship-trained spine surgeons and attempted to evaluate the intraobserver reproducibility and interobserver reliability using plain radiographs and CT images. The aim of this study was to assess the most appropriate imaging modality to which the Powers ratio can be applied in order to diagnose upper cervical spine injury. We hypothesized that the Powers ratio is first a standard and reliable measurement tool for the assessment

of acute upper cervical spine injuries in patients who are suspected of having AOD. We further hypothesized that the use of reformatted CT scans can provide enhanced diagnostic reliability and reproducibility compared with the use of plain radiographs.

Materials and methods

Patients

The present study was carried out using 32 cases with a history of acute cervical spine pain, with lateral plain radiography and reformatted CT scans taken between October 2005 and September 2007 and obtained to evaluate either cervical trauma or degenerative pathology. The cases were selected from a database of two level-one trauma centers. Our series included 22 females and 10 males (age 20–90 years; mean 55 years). There was no difference between the mean Powers ratio based on the patients' age. The protocol for this study was approved by the Institutional Review Board.

Fig. 1 Diagram illustrating the landmarks used to determine the Powers ratio, including the midpoint of anterior arch (A), the opisthion (O), the basion (B), midpoint of posterior arch (C). As mentioned in the text, a BC/AO ratio of >1.0 indicates atlanto-occipital dislocation

Powers ratio

Powers ratio is the ratio of the length of a line extending from the basion to the posterior arch of the atlas divided by that of a line extending from the opisthion to the anterior arch of the atlas (Figs. 1, 2). It is considered abnormal at values greater than one, which was initially described as a measurement of traumatic anterior atlanto-occipital instability, with a normal mean ratio being <0.9 [9].

Procedure

Hard copies of 58 patients with the initial cervical spine radiographs that had been obtained for routine clinical examination were selected for review, including plain radiographs and CT images. From this group, a total 32 patients matched our study criteria. Twenty-six had been excluded either because of the quality of visualization of the radiographic landmarks or due to incomplete radiographic series. The average age of the study group patients was 55 years (range 20–90 years). The selected series had CT images taken at the same time as the radiographs in order to reduce measurement errors from radiographs obtained at different times.

CT images included axial, sagittal, coronal reconstructions. All data regarding the identity of the patient was hidden. The acceptability of each radiograph was determined by each surgeon enrolled as an observer in this study. Five observers independently measured each plain radiograph and CT image using the Powers ratio measurement methodology. The senior author (K.B.W.) instructed five junior authors in the proper measurement techniques prior to the measurements being obtained. The plain radiographs were distributed to the reviewers along with instructions of how to apply the measurement and spreadsheets to record their measurements. Each reviewer interpreted the radiographs independently and was blinded

to how they had been interpreted by the other reviewers. The reviewers were allowed unrestricted time on each plain radiograph, and the measurements were made with a soft pencil by using a plastic millimeter ruler. All marks were completely erased by alcohol after each measurement. Both, the radiographs and the CT images were measured by each reviewer on three separate occasions with at least a 1-month interval between measurements. At the subsequent evaluations, both the plain radiographs and the CT images were presented in a different numeric in order to guard against any recall bias. The reviewers consisted of four spine surgeons and one orthopaedic spine fellow, all with extensive experience in assessment and treatment of cervical trauma.

Statistical analysis

The coefficient of variation (Cv), which is a normalized measure of dispersion of a probability distribution, was chosen for statistical analysis since there were two measurement tools (CT and Plain radiographs) in this study. It is defined as the ratio of the standard deviation (σ) to the mean (μ), $Cv = \sigma/\mu$. This ratio is particularly useful when comparing the variability of two groups and it is generally accepted that a Cv of less than 5% is considered acceptable reproducibility [4].

Any two methods (i.e. CT and Plain radiographs have been selected in this study) that are designed to measure the same instrumental parameter (i.e. Powers ratio) will have good correlation when a set of samples are chosen such that the property to be determined varies a lot between them. Therefore, a Bland-Altman analysis is employed to indicate whether differences between the two measurement methods differ with the quantity being measured [1]. A high correlation for the methods designed to measure the same property is thus simply a sign that one has chosen a wide spread sample. In our study, we performed a comparison of

Fig. 2 a Fracture lucency of right side of C2 is again partially obscured, but persistent mild displacement of the right lateral mass of C2 relative to C1 is unchanged in alignment. **b** No definite change in position or alignment of C2 given this single lateral view of the cervical spine and continued obscuration of the previously described fracture line

the different clinical methods to obtain the Powers ratio to see whether the CT scan and Plain radiographs agree sufficiently and whether the two methods can be used interchangeably. We identified the comparison as determining the limits of agreement, which can be defined with a Bland-Altman appropriated analysis. It implies the plot of the ‘true’ value is not known, but the best estimate for it is the mean of the two measurements. Hence, the Cartesian coordinates of a given sample S with values of S_1 and S_2 can be determined by the two assays is

$$S(x, y) = \left(\frac{S_1 + S_2}{2}, (S_1 - S_2) \right)$$

The limit of agreement is taken as the bias $\pm (1.96 \times \sigma)$, between which 95% of the differences in measurements between the two observers lie. Additionally, the larger the standard deviation the greater is the degree of impression, and the correlation coefficient should not be used.

Results

Agreement on diagnosis

In most cases (88%) there was no evidence of cervical spine fractures, mal-alignment or dislocation on CT scans but 10 of the 32 patients had an abnormal Powers ratio (>1) on plain radiographs. Subsequently, four patients (12%) were diagnosed with C1-2 fractures, retrolisthesis or dislocation but had normal Powers ratio (<1) on both CT image and plain radiographs (Table 1).

With the use of CT scans, there was unanimous agreement for all patients during the three separate measurements performed by the five blinded observers. However, with the use of plain radiographs, there was agreement for only: 21 of the 32 measurements upon the initial reading, 24 of the 32 upon second reading and 25 of the 32 upon the third reading. All observers agreed on the presence of the four fractures during the three readings when the CT images were utilized. However, when plain radiographs were used, only two observers agreed on the presence of the two fractures during the initial reading and there was no agreement upon the second and third readings.

Table 1 Cv mean values for interobserver reliability and intraobserver reproducibility

Observer	CT scans (%)	X-ray (%)
Interobserver reliability	4.31	9.09
Intraobserver reproducibility	4.95	14.17
Highest	9.08	34.3
Lowest	1.39	7.54

The numbers in bold denote statistically significant difference at $P < 0.05$

Intraobserver reproducibility

With respect to intraobserver reproducibility of the five observers’ assessment of Powers ratio, the Cv values ranged from 1.39 to 9.08% using CT scans, with a mean of 4.95%. This corresponds to a level of “acceptable” reproducibility (Table 1). With the use of plain radiographs, the Cv value for intraobserver reproducibility ranged from 7.54 to 34.30%, with a mean of 14.17%; representing “not acceptable” reproducibility.

Interobserver reliability

When CT scans were used to determine the Powers ratio, the mean Cv value for interobserver reliability was 4.31% for the three readings, representing “acceptable” reliability. Using plain radiographs, a Cv mean value of 9.09% was obtained, signifying that there was “not acceptable” agreement (Table 1).

Analysis of agreement between plain radiography and CT

The limits of agreement for plain radiographs and CT scans were -0.8 and 0.89% , respectively (Fig. 3). The two methods produced results that were within the limits of agreement, suggesting that given these limits, the CT and plain radiographic measurements can be used interchangeably.

Discussion

To our knowledge, this is the first study to evaluate the intraobserver reproducibility and interobserver reliability of Powers ratio by a group of observers from different institutions. We found that the interobserver reliability for the plain radiographic measurement of the Powers ratio

Fig. 3 Bland-Altman plot representing the agreement between the two imaging techniques used to measure the Powers ratio in this study. Means of measurements obtained with both methods are averaged and the difference is presented as a function of the mean

was below the acceptable margin of $Cv < 5\%$. Measurements using CT images were more reliable with regard to intra- and interobserver agreement.

Powers ratio, originally reported by Powers in 1979, was initially described as a measurement of traumatic anterior atlanto-occipital instability [9]. Subsequently, methods such as the atlanto-dens interval, Wiesel–Rothman measurement and occiput-atlas angle have been developed for the diagnosis of upper cervical spine injuries [2, 3, 6–8, 10, 11]. Wellborn et al. [11], previously reported on the intraobserver reproducibility and interobserver reliability of several measurements of the atlanto-occipital junction, including Powers ratio, when measured using plain radiographic measurements. They found high reproducibility with coefficients of 0.51 and 0.55, respectively. The coefficient for observer one was, however, only 0.19, raising the possibility of experience as a major variable. The reproducibility of the Powers ratio was poor with the range from 0.05 for observer 2 to 0.33 for a separate observer. Our study was of a much larger group of participants and the coefficient of variation demonstrated more representative clinical significance. This emphasizes the difficulties involved with the use of plain radiographs for making these measurements. For instance, we found that the reproducibility of Cv mean value was up to 34.3%, in other words, the measured Powers ratio using plain radiographs demonstrated very poor intraobserver reproducibility. However, the Bland-Altman analysis showed that the CT and plain radiography were within the -0.8 and 0.89% limits of agreement suggesting that given this tolerance, they can be used interchangeably in clinical practice.

In our study, fellowship-trained spine surgeons demonstrated high reliability and reproducibility only when CT images were used. On plain radiographs, the measurement was not found to be as reproducible and reliable. In addition to its inconclusiveness, plain radiography is somewhat complicated to use as it requires that the observer analyze many more variables than were required by on-site systematic and mechanistic systems. When the Powers ratio was measured on plain radiographs, all readers involved in this study agreed that, because of the complicated topology and many overlapping structures, they had no accurate choice for determining many of the landmarks required for the measurement. Many of the various topological elements such as the opisthion, basion, and mid-point of the posterior arch of the atlas can be visualized more readily with modern imaging techniques (CT) than it could when the Powers ratio was first proposed.

The results of the intraobserver comparison are of particular interest. Six months after their first assessments, the same five spine surgeons, measured the same patients very differently from when they had measured them initially

when plain radiographic films were used (Cv mean value of 14.17%) but consistently when the CT scans were used (Cv of 4.95%). The intraobserver reproducibility among the different readers had significant differences with the highest being 34.30 and 9.08% of Cv mean value and the lowest being 7.54 and 1.39% of Cv mean value using plain radiograph and CT scan, respectively. The high variance between assessments and observers on plain radiograph may be a result of several factors. First, Powers ratio is based on landmarks, such as basion and opisthion, which can be difficult to find in every patient and on every radiograph. Second, their profiles can be altered by slight changes in patient position. Therefore, it is important that exact techniques be used to improve the reproducibility of these landmarks.

This study should be interpreted bearing in mind its potential limitations. It should be noted that the plain radiographic images can vary with patient positioning and CT scans can also vary depending on the parasagittal plane selected in the reformatting. These are factors that can introduce inconsistencies in the data causing greater variance and poorer repeatability and reliability. However, we tried to control for these inconsistencies by including images obtained with radiological protocols that are standardized among our institutions.

Despite the higher variance and inferior repeatability and reliability of the plain radiographic measurements, as determined by the Cv , the Bland-Altman plot showed that, given the -0.8 and 0.89% limits of agreement, the two techniques have sufficient agreement to the point that they can be used interchangeably [11]. Therefore, plain radiography may still be valuable in the initial management of upper cervical trauma with suspected atlanto-occipital dislocation considering its accessibility, speed, lower radiation and cost.

Based on our data, we believe that the Powers ratio method is not sufficiently reliable when it is obtained only using plain radiographs. However, although it was shown to have inferior repeatability and reliability to that of CT, the Bland-Altman plot demonstrated sufficient agreement between the two techniques within the -0.8 and 0.89% limits of agreement, suggesting that given this tolerance for agreement they could be used interchangeably in the assessment of atlanto-occipital junction.

References

1. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476):307–310
2. Dziurzynski K, Anderson PA, Bean DB, Choi J, Levenson GE, Marin RL, Resnick DK (2005) A blinded assessment of

- radiographic criteria for atlanto-occipital dislocation. *Spine* 30(12):1427–1432. doi:[10.1097/01.brs.0000166524.88394.b3](https://doi.org/10.1097/01.brs.0000166524.88394.b3)
3. Fehlings MG et al (2006) Interobserver and intraobserver reliability of maximum canal compromise and spinal cord compression for evaluation of acute traumatic cervical spinal cord injury. *Spine* 31(15):1719–1725. doi:[10.1097/01.brs.0000224164.43912.e6](https://doi.org/10.1097/01.brs.0000224164.43912.e6)
 4. Fleiss JL (1981) Statistical methods for raters and proportions, 2nd edn. Wiley, New York, pp 212–236
 5. Gaa J, Deininger HK (1989) Traumatic atlanto-occipital dislocation in conventional x-ray diagnosis. *Radiologe* 29(7):354–358
 6. Hamai S, Harimaya K, Maeda T, Hosokawa A, Shida J, Iwamoto Y (2006) Traumatic atlanto-occipital dislocation with atlantoaxial subluxation. *Spine* 31(13):E421–E424. doi:[10.1097/01.brs.0000220224.01886.b3](https://doi.org/10.1097/01.brs.0000220224.01886.b3)
 7. Karol LA, Sheffield EG, Crawford K, Moody MK, Browne RH (1996) Reproducibility in the measurement of atlanto-occipital instability in children with Down syndrome. *Spine* 21(21):2463–2467. doi:[10.1097/00007632-199611010-00010](https://doi.org/10.1097/00007632-199611010-00010) (discussion 2468)
 8. Kuzma BB, Goodman JM (1997) Diagnosis of atlanto-occipital dislocation. *Surg Neurol* 48(4):418–419. doi:[10.1016/S0090-3019\(97\)00396-0](https://doi.org/10.1016/S0090-3019(97)00396-0)
 9. Powers B, Miller MD, Kramer RS, Martinez S, Gehweiler JA Jr (1979) Traumatic anterior atlanto-occipital dislocation. *Neurosurgery* 4(1):12–17. doi:[10.1097/00006123-197901000-00004](https://doi.org/10.1097/00006123-197901000-00004)
 10. Saeheng S, Phuenpathom N (2001) Traumatic occipitotlantal dislocation. *Surg Neurol* 55(1):35–40. doi:[10.1016/S0090-3019\(00\)00350-5](https://doi.org/10.1016/S0090-3019(00)00350-5) (discussion 40)
 11. Wellborn CC, Sturm PF, Hatch RS, Bomze SR, Jablonski K (2000) Intraobserver reproducibility and interobserver reliability of cervical spine measurements. *J Pediatr Orthop* 20(1):66–70. doi:[10.1097/00004694-200001000-00015](https://doi.org/10.1097/00004694-200001000-00015)