

Clinical Article

Analysis of Measurement Accuracy for Craniovertebral Junction Pathology : Most Reliable Method for Cephalometric Analysis

Ho Jin Lee, M.D., Jae Taek Hong, M.D., Ph.D., Il Sup Kim, M.D., Ph.D., Jae Yeol Kwon, M.D., Sang Won Lee, M.D., Ph.D.

Department of Neurosurgery, St. Vincent's Hospital, The Catholic University of Korea, Suwon, Korea

Objective : This study was designed to determine the most reliable cephalometric measurement technique in the normal population and patients with basilar invagination (BI).

Methods : Twenty-two lateral radiographs of BI patients and 25 lateral cervical radiographs of the age, sex-matched normal population were selected and measured on two separate occasions by three spine surgeons using six different measurements. Statistical analysis including intraclass correlation coefficient (ICC) was carried out using the SPSS software (V. 12.0).

Results : Redlund-Johnell and Modified (M)-Ranawat had a highest ICC score in both the normal and BI groups in the inter-observer study. The M-Ranawat method (0.83) had a highest ICC score in the normal group, and the Redlund-Johnell method (0.80) had a highest ICC score in the BI group in the intra-observer test. The McGregor line had a lowest ICC score and a poor ICC grade in both groups in the intra-observer study. Generally, the measurement method using the odontoid process did not produce consistent results due to inter and intra-observer differences in determining the position of the odontoid tip. Opisthion and caudal point of the occipital midline curve are somewhat ambiguous landmarks, which induce variable ICC scores.

Conclusion : On the contrary to other studies, Ranawat method had a lower ICC score in the inter-observer study. C2 end-plate and C1 arch can be the most reliable anatomical landmarks.

Key Words : Cephalometric measurement · Basilar invagination · Odontoid process · Opisthion · C2 end plate · C1 arch.

INTRODUCTION

Basilar invagination (BI) has many synonyms like cranial settling, vertical settling, vertical atlanto-axial subluxation and atlanto-axial impaction, and is defined as superior migration of the odontoid tip into the foramen magnum, leading to compression of the brainstem.

BI prevalence may be not as rare as we thought before, although the prevalence of BI is less common than other spine diseases in general. Mikulowski et al.⁹⁾ showed that 11 (10%) patients had unrecognized cord compression, which can be the cause of death in rheumatoid arthritis patients, and we also found 20 (8.2%) patients with BI in 243 rheumatoid arthritis patients who were admitted to our hospital¹¹⁾. The early diagnosis of BI was typically conducted by standard radiographs in

most cases. However, no one-single method has been recommended to properly diagnose BI, due to the overlying structures on lateral plain radiographs. Ambiguous landmarks lead to low reliability or consistency for confirming BI. Accordingly, we need to establish a more reliable and consistent method for early diagnosis of BI.

The purpose of this study was to determine the inter-observer reliability and intra-observer repeatability of the methods for assessing the most appropriate measurement.

MATERIALS AND METHODS

We chose 22 BI patients (female 16, male 6), who were confirmed by MRI or CT as having BI. We selected other 25 patients (female 17, male 8) matching in age and sex as normal

• Received : January 21, 2013 • Revised : June 26, 2013 • Accepted : September 30, 2013

• Address for reprints : Jae Taek Hong, M.D., Ph.D.

Department of Neurosurgery, St. Vincent's Hospital, The Catholic University of Korea, 93 Jungbu-daero, Paldal-gu, Suwon 442-723, Korea
Tel : +82-31-249-7196, Fax : +82-31-249-5208, E-mail : jatagi15@paran.com

• This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

controls, using the electrical medical recording system. Cervical lateral radiographs of a total of 47 patients were selected for review. This study was done from June 2011 to August 2011. The average age of the study group was 67.2 (BI group) years and 67.6 (control group) years.

The local research ethics committee waived the need for formal ethics approval for this retrospective study.

Radiographic analysis

Only one cervical lateral standard radiograph per patient was saved in Picture Archiving Communication System (PACS) by a senior neurosurgeon in a randomized order. Thus, all data of the BI group and the control group were arranged in the same data folder. Also, all data regarding the identity of the patients were

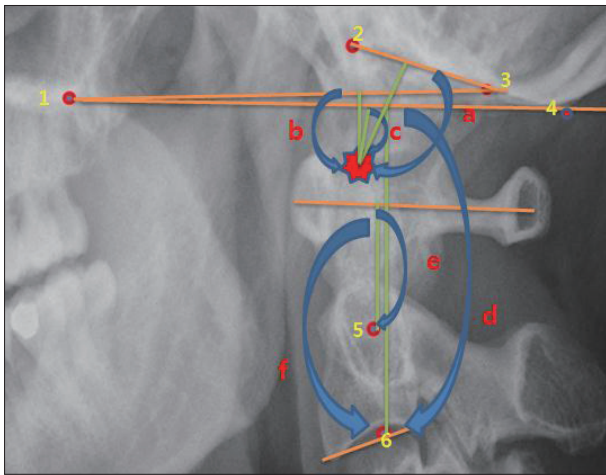


Fig. 1. Relevant landmarks and six-different measurements. 1 : Hard palate, 2 : Basion, 3 : Opisthion, 4 : The most caudal point on the midline occipital curve, 5 : Center of the second cervical pedicle, 6 : Midpoint of the caudal margin of the second cervical vertebra body, a : McRae line, b : Chamberlain line, c : McGregor line, d : Redlund-Johnell method, e : Ranawat method, f : Modified-Ranawat method, Asterisk : odontoid tip.

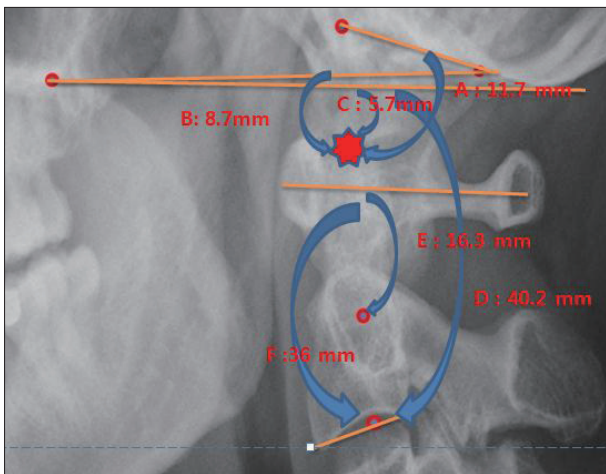


Fig. 2. Measuring the six-differents methods in normal patient. A : McRae method, B : Chamberlain method, C : McGregor method, D : Redlund-Johnell method, E : Ranawat method, F : Modified-Ranawat method, Asterisk : odontoid tip.

blocked out. The most true-lateral of cervical spine radiographs was selected among many follow-up radiographs to reduce the super-imposition factor. The lateral cervical radiographs were evaluated by three blinded neurosurgeons. Three observers independently performed the measurement and checked the relevant anatomical landmarks and measuring technique before measuring without prior knowledge of the patients.

Time was not restricted in measuring the radiograph, and the measurement were done with an electric caliper in PACS. The radiographs were measured by each observer on two separate occasions with at least a two-month interval between measurements for removing the after-image effect. After-image effect can create a bias in the intra-observer measurement results, in particular. At the second evaluation, the radiographs were saved with a different numeric order to guard against any recall-bias.

The results were tabulated for each observer and intra-observer as well as intra-observer agreement was assessed using the intraclass correlation coefficient (ICC) test. The BI group and the control group were analyzed separately.

Measurement

We used six different measurements for the study and the measuring technique was followed as described by the original author's paper. McGregor line, McRae line, Chamberlain line, Ranawat method, Modified (M)-Ranawat method, and Redlund-Johnell method were included in this study. The following section explains the six different measurements in detail (Fig. 1-3).

McGregor line : A line is drawn from the posterosuperior aspect of the hard palate to the most caudal point on the midline occipital curve. Protrusion of the odontoid-tip above this line was represented with negative number⁷⁾.

McRae line : A line is drawn across the foramen magnum from the basion to the opisthion. Protrusion of the odontoid-tip above this line was represented with a negative number⁸⁾.

Chamberlain line : A line is drawn from the posterior edge of the hard palate to the opisthion. Protrusion of the odontoid-tip above this line was represented with a negative number¹⁾.

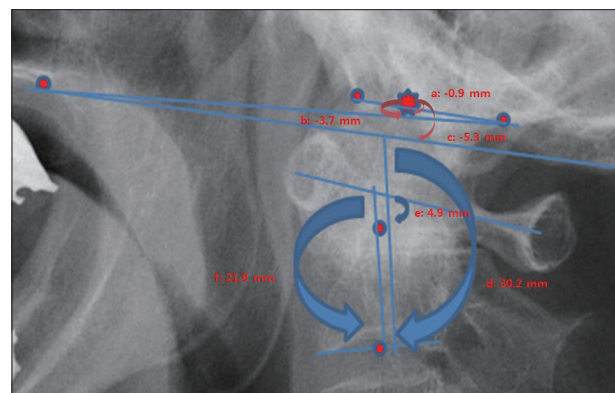


Fig. 3. Measuring the six-differents methods in basilar invagination patient. a : McRae method, b : Chamberlain method, c : McGregor method, d : Redlund-Johnell method, e : Ranawat method, f : Modified-Ranawat method, Asterisk : odontoid tip.

Ranawat criterion : The distance between the center of the second cervical pedicle and the transverse axis of the atlas is measured along the axis of the odontoid process¹².

Modified (M)-Ranawat criterion : The distance between the midpoint of the base of C2 end-plate and a line from the center of the anterior arch of C1 to the center of the posterior arch⁶.

Redlund-Johnell method : The distance between the McGregor line and the midpoint of the caudal margin of the second cervical vertebra body is measured¹³.

Statistical analysis

Reliability was examined using ICC score and their 95% confidence intervals. A *p* value smaller than 0.05 was considered significant. This analysis reflects agreement on the repeated measurements regardless of who performed the measurement. The ICC ranged from 0 to 1, where 0 represented no agreement and 1 perfect agreement. Data analysis was carried out using SPSS software (V. 12.0).

RESULTS

Interobserver reliability

ICC score of all measurements was higher in the normal group than the BI pathology group, except in the chamberlain line method (Fig. 4). Redlund-Johnell and M-Ranawat had a highest ICC score in both the normal and BI groups. McRae line (0.21) had a lowest ICC score in the normal group, and the Ranawat method (0.18) had a lowest ICC value in the BI pathology group. McRae and Ranawat methods had poor grades

of ICC in both groups.

Intraobserver reliability

ICC score of all measurements was higher in the normal group than BI group, except in the Redlund-Johnell method (Fig. 5). The M-ranawat method (0.83) had a highest ICC score in the normal group, and Redlund-Johennll method (0.80) had a highest ICC score in the BI group. McGregor line has a lowest ICC score and a poor ICC grade in both groups.

DISCUSSION

Having many methods for the diagnosis of BI can imply that it is very difficult to choose just one particular method in clinical circumstances. These measurements can show variable results, due to multiple reasons. First, anatomical landmark may be ambiguous, thus leading the interpreter to measure different results. Second, measurement error can be made by the interpreter himself or on the radiographs. The lack of confidence in anatomic landmarks can make unreliable results and it is hard to obtain the absolute true-lateral radiographs in every patients.

Variation in measurement may lead to a different type of treatment. Therefore it is very important for us to determine how reliable, reproducible these measurements are. We verified the reliability and reproducibility among various measurement techniques with inter-observer and intra-observer correlation studies. This study was performed using six-different methods, excluding Clark station², Kauppi et al.⁴, and Wackenheim line method, because these methods can not represent numerical

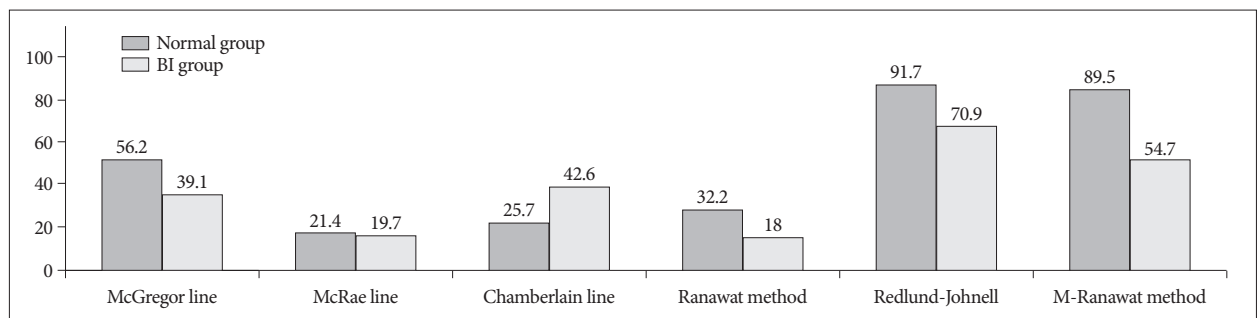


Fig. 4. This graphs show the intraclass correlation coefficient score (original ICC score times one hundred) with inter-observer study. BI : basilar invagination.

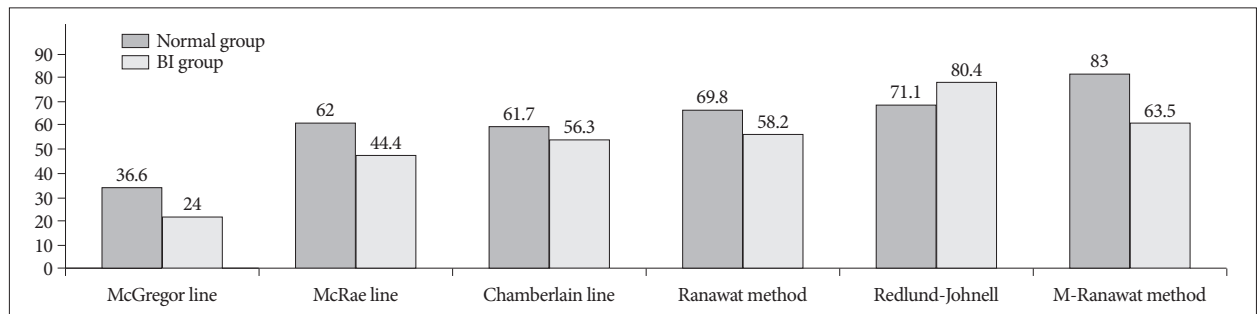


Fig. 5. This graphs shows the intraclass correlation coefficient score (original ICC score times one hundred) with intra-observer study. BI : basilar invagination.

value, so it will be useless in the present inter-observer and intra-observer reliability test. Also, Wackenheim line has been shown to have low specificity in many reports¹⁴. Yune et al. revealed that the reason that dorsal surface of clivus is rarely a straight line, unlike its appearance on radiographs⁶.

Some presumptions were made before conducting this study. First, ICC score will be higher in the normal group than the BI group, because the normal group has relatively precise anatomic landmarks than the BI groups. Second, intra-observer correlation may gain the upper hand than inter-observer correlation. Third, shared anatomic landmarks between many diagnostic methods may be the key to make the similar pattern of results. For example, the C1 arch is the key between Ranawat and M-Ranawat and the Opisthion is the key between McRae and Chamberlain. The caudal point of the occipital curve is the key between McGregor and Redlund-Johnell, the midpoint of the base of C2 endplate is the key between Redlund-Johnell and M-Ranawat method. The odontoid tip is the key in McGregor, McRae and Chamberlain line method.

Generally, the intra-observer correlation value was higher than inter-observer correlation in our study, which was consistent with our assumption. Intra-observer reproducibility was related to consistency for measuring pattern in each observer in determining the anatomic landmarks and using the PACS system. Therefore, all observers had their own specific measuring pattern, although it is very difficult to identify the most correct pattern. If we can re-examine the measuring process with all examined radiographs, which remained with the electrical trace, it will be a good opportunity to increase the inter-observer correlation and reduce the error.

Inter-observer reliability

Odontoid-tip based measurement (McGregor line, McRae line, Chamberlain line) had a low ICC score than other measurements with the inter-observer test. Many reports showed similar results like the results of the present study, stating the difficulty in identifying the odontoid-tip¹⁴. The odontoid-tip is not clearly visible on standard radiographs, especially with old age or rheumatoid arthritis patients, due to erosion, overlying mastoid process and osteoporosis. We conducted bone densitometry and the mean t-score was -2.27 (range, -3.9 - -0.7) in the BI group and the mean value of t-score was -1.49 (range, -2.3 - -0.4) in the normal group. The mean age was 68.8 years in all 47 patients. More severe osteoporotic patterns were confirmed in the BI group, which can induce the low ICC score in the BI group than the normal group.

The opisthion may be the ambiguous landmark in the inter-observer test, and this reason can be explained by the following clue. First, the ICC score was reversed only in the chamberlain line method between the BI group (42.6) and the normal group (25.7). Second, McRae has a lowest ICC score in both study groups (except in the BI group). A super-imposition factor, induced by relative globular form of skull base, may be the main

reason for the above results. We also must consider the basion as an important attributable factor in lowering the ICC score in the McRae line method³.

Many previous studies showed that the Ranawat method has good sensitivity and may be the one of the best diagnostic tools^{5,14}. However, the Ranawat method showed a lower ICC score (the fourth position in normal, the last position in BI group) in our study than expected. Riew et al.¹⁴ showed that combination of Clark station, Redlund-Johnell and Ranawat method gave high sensitivity (94%) and negative predictive value (91%). If we assume that the C1 arch is a good landmark by comparing the high ICC score in the Redlund-Johnell and M-Ranawat methods, the center of the second cervical pedicle may be the ambiguous landmark. A great difference in the ICC score between the Ranawat and M-Ranawat methods can also represent the reason. C1-2 facet joint destruction was one of the pathophysiologies in basilar invagination and osteoporotic change of the C2 pedicle may be the reason for the results. In addition, the cortical margin of the C2 pedicle was not in an absolute globular form, which can make the observer difficult to decide the exact center point.

Redlund-Johnell and M-Ranawat has a highest ICC score in both groups. Therefore, the midpoint of the caudal margin of the second cervical vertebra body may be the most reliable landmark for the diagnosis of BI pathology, as mentioned above. Vertebral body has relative a plane figure anatomically than other structures and there was no interfering bony structure near the surrounding. In addition, the bony erosion is rare in vertebral body. The Redlund-Johnell, Ranawat method and M-Ranawat method are measures of the spatial relationship between C1 and C2 rather than the more critical occiput-C2 relationship⁶. Basically, basilar invagination pathophysiologic consequence was mainly induced by C1-2 articulation than occiput-C1¹⁰.

Intra-observer reliability

McGregor line was the lowest in ICC score in the intra-observer test. Only the McGregor line has showed the decreasing pattern in the ICC score, between the inter-observer and the intra-observer test. Redlund-Johnell method also has the reversed ICC score between inter-observer (91.7) and intra-observer (71.1) test in normal group. These above findings can imply that the caudal point on the midline occipital curve may also be an ambiguous landmark. The Ranawat method had a relatively higher ICC score than the inter-observer test. Although it is difficult to assure exact the reason, different measuring patterns may exist by individuals, in a similar manner as individual variations in the locations of the center of the second cervical pedicle, and parallel to the axis odontoid-process.

McGregor test has a lowest ICC score compared to other odontoid-tip based methods in our study. This phenomenon can be explained with the shortest distance used in the measurement. Anatomically, McGregor line has the nearest position to the odontoid-tip compared to other methods, which can

induce difficulty in the precise of the electric caliper. Unlike pencils, electrical caliper is operated by cursor-controlled by mouse and magnifying degree is not fixed and strictly regulated. The less the short distance measured with inaccuracy, the less the correlation score we obtain.

Redlund-Johnell and M-Ranawat methods were shown to have the highest ICC score in both groups, as with the inter-observer study.

Limitation and interpretation

All three observers differed in medical standing, one was a senior spine-neurosurgeon (highly experienced) and the others were junior spine-neurosurgeon (less experienced). The observer-bias could be included, even though three observers tried to reduce the problem with prior consent about anatomical landmarks and measuring technique considerably. As mentioned before, consensus on how to use the software system (PACS) was strictly regulated in this study. However, the degree of magnification and the pattern of using the electric caliper were variable, which can attribute to some degree of measuring bias. We also could not hold a conviction with the standardization for neck position, thus occipito-cervical angle and the degree of flexion or extension can be variable in each patient. The degree of measurement in each diagnostic method can be influenced by the neck position and occipito-cervical angle.

Intra- as well as inter-observer reliability are connected to the concept of consistency, which is defined as the agreement of two quantitative measurements where neither one is assumed 'correct'. Therefore, our results may not show the correctness of the methods exactly, but provided some positive contributions. First, these results can open our eyes to the fact that we always must consider the lack of stability in our measuring technique. Second, ICC score may suggest the most reliable anatomic landmark, which may help us to find the best combination of methods, so that not to miss the basilar invagination. Although our study revealed that odontoid-tip based measurements have low ICC score, these measurement can be useful in a CT based-study, which can show the odontoid-tip more accurately. Opisthion and caudal point of the midline occipital curve based measurements can be also significant with efforts to obtain the true-lateral radiographs.

CONCLUSION

Ranawat method is rather variable between inter-observation

and intra-observation. Thus, the center of the C2 pedicle may no longer be a reliable anatomical landmark, which is a different conclusion from other previous studies. Odontoid-tip based method is not a reliable method as revealed by many previous reports. Redlund-Johnell and M-Ranawat are the most reliable methods, thus C1 arch and C2 endplate may be the most reliable anatomical landmarks.

References

1. Chamberlain WE : Basilar Impression (Platybasia) : A Bizarre Developmental Anomaly of the Occipital Bone and Upper Cervical Spine with Striking and Misleading Neurologic Manifestations. *Yale J Biol Med* 11 : 487-496, 1939
2. Clark CR, Goetz DD, Menezes AH : Arthrodesis of the cervical spine in rheumatoid arthritis. *J Bone Joint Surg Am* 71 : 381-392, 1989
3. Dziurzynski K, Anderson PA, Bean DB, Choi J, Levenson GE, Marin RL, et al. : A blinded assessment of radiographic criteria for atlanto-occipital dislocation. *Spine* 30 : 1427-1432, 2005
4. Kauppi M, Sakaguchi M, Kontinen YT, Hamalainen M : A new method of screening for vertical atlantoaxial dislocation. *J Rheumatol* 17 : 167-172, 1990
5. Kawaida H, Sakou T, Morizono Y : Vertical settling in rheumatoid arthritis. Diagnostic value of the Ranawat and Redlund-Johnell methods. *Clinical Clin Orthop Relat Res* 239 : 128-135, 1989
6. Kwong Y, Rao N, Latief K : Craniometric measurements in the assessment of craniovertebral settling : are they still relevant in the age of cross-sectional imaging? *AJR Am J Roentgenol* 196 : W421-W425, 2011
7. Mc GM : The significance of certain measurements of the skull in the diagnosis of basilar impression. *Br J Radiol* 21 : 171-181, 1948
8. McRae DL, Barnum AS : Occipitalization of the atlas. *Am J Roentgenol Radium Ther Nucl Med* 70 : 23-46, 1953
9. Mikulowski P, Wollheim FA, Rotmil P, Olsen I : Sudden death in rheumatoid arthritis with atlanto-axial dislocation. *Acta Med Scand* 198 : 445-451, 1975
10. Nagayoshi R, Ijiri K, Takenouchi T, Taketomi E, Sakakima H, Komiya S : Evaluation of occipitocervical subluxation in rheumatoid arthritis patients, using coronal-view reconstructive computed tomography. *Spine* 34 : E879-E881, 2009
11. Parish DC, Clark JA, Liebowitz SM, Hicks WC : Sudden death in rheumatoid arthritis from vertical subluxation of the odontoid process. *J Natl Med Assoc* 82 : 297-299, 302-304, 1990
12. Ranawat CS, O'Leary P, Pellicci P, Tsairis P, Marchisello P, Dorr L : Cervical spine fusion in rheumatoid arthritis. *J Bone Joint Surg Am* 61 : 1003-1010, 1979
13. Redlund-Johnell I, Pettersson H : Radiographic measurements of the cranio-vertebral region. Designed for evaluation of abnormalities in rheumatoid arthritis. *Acta Radiol Diagn (Stockh)* 25 : 23-28, 1984
14. Riew KD, Hilibrand AS, Palumbo MA, Sethi N, Bohlman HH : Diagnosing basilar invagination in the rheumatoid patient. The reliability of radiographic criteria. *J Bone Joint Surg Am* 83 : 194-200, 2001