



Original article

Agreement on grading of normal clivus using magnetic resonance imaging among radiologists

Sujitra Tinnut^a, Waneerat Galassi^a, Nongluk Oilmungmool^a, Thanin Chattrapiban^{b,*}

^a Department of Radiology, Faculty of Medicine, Naresuan University, 99 Moo 9, Phitsanulok-Nakhon Sawan Road, Tha Pho, Mueang Phitsanulok, Phitsanulok 65000, Thailand

^b Clinical Epidemiology & Clinical statistics Division, Department of Family Medicine, Faculty of Medicine, Naresuan University, 99 Moo 9, Phitsanulok-Nakhon Sawan Road, Tha Pho, Mueang Phitsanulok, Phitsanulok 65000, Thailand

ARTICLE INFO

Keywords:

Clivus
Marrow conversion
MRI
Agreement
Kappa statistic

ABSTRACT

Purpose: The present study was aimed to evaluate the agreement on grading normal clivus on MRI among radiologists.

Methods: A retrospective study included patients who underwent MRI brain during January 1, 2015 to October 31, 2019. Two hundred forty-four patients who had no marrow pathology on MRI were included and divided into 8 age groups by decades. Three radiologists independently reviewed the signal intensity of clivus in mid sagittal T1-weighted image. The signal intensity was classified into three grades (Grade I-III). Fleiss' kappa coefficients (k) were calculated to assess interrater agreement.

Results: Of 244 patients, there were 123 (50.4%) males and 121 (49.6%) females. Age ranged from 1 to 79 years old. Clivus Grade II was more frequently reported (> 50%) by radiologists. The agreement (kappa) among all three radiologists on evaluation of clivus irrespective of the grading equals to 0.67 (95%CI: 0.60–0.74). In stratified analyses by the grade of clivus, the kappa values for Grade I to III and were 0.73, 0.62, and 0.69 respectively.

Conclusion: Interrater agreement of MRI evaluation of normal clivus among radiologists was good. The visual grading criteria to classify the clivus is sufficient to distinguish the marrow maturation. However, the consensus reading should be made whenever normal clivus Grade II is read.

1. Introduction

Bone marrow in the neonate is almost entirely red marrow. A progressive conversion to fatty marrow occurs generally from appendicular to axial skeleton. Therefore, marrow conversion of clivus is slow and proceeds throughout life [1]. Clivus is a midline skull base and well assessed on the sagittal T1-weighted image of routine brain magnetic resonance imaging (MRI).

MRI is the modality of choice for evaluating bone marrow. Age profoundly affects marrow signal intensity (SI) on MRI depending on the relative amount of protein, water, fat, and cells within the marrow. Signal intensity increases as the hematopoietic cells become predominant, and consequently, trabeculae sparse in the marrow cavity. Knowledge of the orderly pattern with age is valuable to obtain baseline for defining potential bone marrow abnormality and can confidently

evaluate physiologic marrow conversion.

To date, two grading systems for clivus classification on MRI were published based on the comparison of SI of clivus relative to pons [2,3]. Both systems are similarly used to classify clivus into three grades (Grade I, II, and III), but the criteria are different. The first system mainly focuses on gross visualization of the signal heterogeneity which is simply classified into uniform and multiform appearances [3–5]. The second one in contrast uses the amount of percentage of signal heterogeneity rather than using gross visualization as in the first system [2,6]. In our hypothesis, we believe that the second grading system is more accurate and more efficient assessment compared to the first system. However, there is no evidence whether using this system would yield consistent readings among our radiologist team.

Many previous studies have reported physiologic signal changes of clivus by MRI, but the assessment of the agreement on grading the clivus

* Corresponding author.

E-mail addresses: sujitrati@nu.ac.th (S. Tinnut), gwaneerat@gmail.com (W. Galassi), oy_ing@hotmail.com (N. Oilmungmool), thaninc@nu.ac.th (T. Chattrapiban).

<https://doi.org/10.1016/j.ejro.2022.100395>

Received 16 October 2021; Accepted 6 January 2022

Available online 13 January 2022

2352-0477/© 2022 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

among radiologists has not been evaluated [2–6]. Assessing the agreement could inform the use of current grading system of clival physiologic changes and might be indicative of improving the current grading system. Therefore, the purpose of our study was to assess whether the grading of clivus based on MRI is well agreed among radiologists.

2. Material and methods

2.1. Patients

This retrospective study was approved by the ethics committee. We included 1717 patients aged more than 1 year who underwent brain MRI in our radiology department between January 1, 2015 to October 31, 2019. A sagittal T1-weighted sequence of MRI brain has been performed in all patients. Patients with the following conditions were excluded from the study (1) age more than 79 years, (2) having abnormalities of pons, (3) having history of previous chemotherapy or radiation, and (4) having systemic diseases such as lymphoproliferative disorder, sickle cell anemia, anaplastic anemia, osteomyelitis, and systemic lupus erythematosus (SLE). We excluded 120 patients who met the exclusion criteria. Then, we performed a stratified random sampling by computer-generated random numbers. The strata of patient's age was divided into 8 age groups; (1–9 years, 10–19 years, 20–29 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, and 70–79 years). Approximately 30 patients in each age group were randomly chosen by computer and we had a total number of 244 patients for our analysis.

2.2. Imaging analysis

The patients that met the inclusion criteria were scanned on a 1.5 Tesla scanner (Philips Ingenia, Philips Medical Systems, Best, the Netherlands) by using a head array coil. Sagittal spin-echo T1-weighted sequence was acquired with the following parameters: TR = 500–600 msec, TE = 11–17 msec, FOV 21 cm, slice thickness 5 mm, number of signal acquisitions = 1–2, and a 192 × 192 matrix size. The mid sagittal image was chosen for assessment as it can identify midline structures such as the clivus, genu of corpus callosum, pons, and the fourth ventricle on the same plane (Fig. 1). Three radiologists having different

levels of experiences (two neuroradiologists (A and C, with 16 years and 1 year of experiences respectively) and one pediatric radiologist (B with 6 years of experiences)) independently reviewed the images of all included patients. In this study, we used the predetermined criteria by Kimura et al. [2] which graded clivus according to its SI as compared to pons. The clivus was divided into three grades: Grade I, II, and III (Fig. 2). Grade I was defined as a predominantly low SI in more than 50% of the clivus. For Grade II, the visualization of low SI portion occupied between 20% and 50% of the clivus area. Grade III was recognized as dominantly high SI which represented low SI occupying less than 20% of the clivus. Three radiologists operated the training session before the study had been carrying out such that the training would gain more confident. The MRI used for the training were not included in the study. The brain images were blinded to the radiologists with respect to the patients' name, age, and hospital identification. After each radiologist had given a grade of the clivus to all images, the consensus grading was made. In case of disagreement between the grades, the majority of vote for such grade determined the consensus grading among radiologists.

2.3. Statistical analyses

All statistical analyses were performed in RStudio version 1.0.136 and Stata version 12.1. Descriptive statistics for numerical data were presented as mean and standard deviation (SD), and categorical data were presented as frequencies and percentages. We used “IRR” package in R to assess the interrater agreement. The agreement of grading the clivus (category variable; Grade I, II, and III) between two radiologists was presented by the Cohen's kappa coefficients (k) that is corrected for the chance agreement among radiologists. Since the grades of the clivus are more than two levels, we used the weighted Cohen's kappa to correct the partial agreement among radiologists. The linear weights are defined. For the assessment of agreement among three radiologists, we corrected for the chance agreement with Fleiss's formula [7]. Therefore, the agreement of grading the clivus by three radiologists was presented by Fleiss's kappa coefficients. We classified the strength of agreement based on Landis and Koch criteria in which the values of k were categorized as follow; < 0.20 (Poor), 0.21–0.40 (Fair), 0.41–0.60 (Moderate), 0.61–0.80 (Good), and 0.81–1.00 (Very good) [7]. The interval estimation for the agreement between two radiologists was calculated and presented as 95% confidence intervals (CI). In addition, we performed a bootstrap resampling method to calculate 95% CI for Fleiss kappa coefficients using the command in the same IRR package in R.

3. Results

Of 244 patients with normal clivus between 1 and 79 years old, there were 123 (50.4%) males and 121 (49.6%) females. Grade II was identified more than 50% of patients by each radiologist. Grade I was read at the highest percentage (22%, $N = 55$) by C radiologist than the others (less than 16%). Grade III was read about 20% of patients for each radiologist. By consensus grading, Grade I was identified in 43 cases (17.6%), Grade II in 147 cases (60.3%), and Grade III in 53 cases (21.7%) (Table 1).

3.1. Assessment of the agreement

Fleiss kappa among all three radiologists on evaluation of clivus irrespective of the grading equals to 0.67 (95%CI:0.60–0.74). The kappa for A and B radiologists showed the highest ($k = 0.73$, 95%CI: 0.65–0.81) while that for A and C radiologists was the lowest ($k = 0.69$, 95%CI: 0.61–0.77) (Table 2). In the stratified analyses by the grade of clivus (Table 3), Grade I had the highest total observed agreement among radiologists (92.1%) and also showed the highest kappa ($k = 0.73$), while Grade II showed the lowest total observed agreement (81.7%) and kappa ($K=0.62$). Grade III has the total observed agreement



Fig. 1. Mid sagittal T1-weighted image shows genu of corpus callosum (asterisk), clivus (arrow), pons, and the fourth ventricle (arrowhead) in the same plane.



Fig. 2. MRI shows bone marrow appearance of clivus Grade I (a), Grade II (b), and Grade III (c).

Table 1

Characteristics of patients and the grading of clivus by radiologists (n = 244).

| Characteristics | n (%) |
|--------------------------------|------------|
| Gender | |
| Male | 123 (50.4) |
| Female | 121 (49.6) |
| Age (years) (Min - Max) | (1 - 79) |
| Reading by A | |
| Grade I | 37 (15.2) |
| Grade II | 157 (64.3) |
| Grade III | 50 (20.5) |
| Reading by B | |
| Grade I | 39 (16.0) |
| Grade II | 154 (63.1) |
| Grade III | 51 (20.9) |
| Reading by C | |
| Grade I | 55 (22.5) |
| Grade II | 132 (54.1) |
| Grade III | 57 (23.4) |
| Consensus grading | |
| Grade I | 43 (17.6) |
| Grade II | 147 (60.3) |
| Grade III | 54 (22.1) |

Table 2

Kappa statistics between two radiologists and among all radiologists.

| Radiologists | Kappa ^a | 95% CI | P-value |
|------------------------|--------------------|-----------|---------|
| A vs. B | 0.73 | 0.65–0.81 | < 0.001 |
| A vs. C | 0.69 | 0.61–0.77 | < 0.001 |
| B vs. C | 0.71 | 0.64–0.80 | < 0.001 |
| All three radiologists | 0.67 | 0.60–0.74 | NA |

^a Fleiss's kappa statistics and bootstrapping for 95% Confident interval (CI)

89.6% and has kappa 0.69.

In additional analyses, the grading of clivus from each radiologist was well agreed with the consensus grading (Table S1). For example, Grade I and III were correctly read by each radiologist more than 80% of patients when comparing to the consensus grading. Fleiss's kappa on grading of clivus by radiologists was different by each age subgroup (Table S2). High kappa values were observed in the age of 1–10 years ($k = 0.71$), 11–20 years ($k = 0.85$), and age > 60 years ($k = 0.72$), whereas low kappa values were observed in the age of 21–40 years ($k = 0.57$) and 41–60 years ($k = 0.49$).

4. Discussion

Our study showed that the agreement represented by kappa statistic among three radiologists on reading MRI of clivus equals to 0.67 which was in a 'good' category based on the classification system [7]. In the

Table 3

Agreement of clivus among three radiologists according to the grades.

| Grading | Radiologists | K ^a /Agreement ^b (%) | | |
|-----------|--------------|--|-------|-------|
| | | A | B | C |
| Grade I | A | – | 94.3% | 91.0% |
| | B | 0.78 | – | 91.0% |
| | C | 0.70 | 0.71 | – |
| | Total | 0.73 / 92.1% | | |
| Grade II | A | – | 84.0% | 79.9% |
| | B | 0.65 | – | 81.1% |
| | C | 0.59 | 0.61 | – |
| | Total | 0.62 / 81.7% | | |
| Grade III | A | – | 89.7% | 88.9% |
| | B | 0.68 | – | 90.2% |
| | C | 0.67 | 0.71 | – |
| | Total | 0.69 / 89.6% | | |

^a Cohen kappa.

^b Overall observed agreement.

analysis of agreement stratified by the grades of clivus, the results showed that the agreement among radiologists was higher in Grade I ($k = 0.73$) and Grade III ($k = 0.69$) than in Grade II ($k = 0.62$).

The overall agreement among three radiologists on reading clivus grade of brain MRI was good. However, it is evident that the overall agreement was attenuated by the reading Grade II of clivus. Some factors could affect our results. Using the visual grading system of clivus on brain MRI is one important factor which could influence the overall agreement because the virtual grading system used in our study relies on the visual judgment of radiologists to classify bone marrow signal alteration [2,6]. Reading Grade I and III have not much been affected by the grading system since it is simply based on predominant SI of the clivus i.e., low SI (represented by Grade I) and high SI (represented by Grade III). In contrast, Grade II is more challenging to read since it is defined by low SI ranged from 20% to 50% of clivus area; therefore, reading Grade II is more variable among radiologists. Our finding might reflect that the current grading system of MRI clivus used in our setting is helpful for grading of clivus on MRI, especially for differentiating Grade I and III, but not Grade II. In additional analysis (Table S2), we showed that age of study participants is the other factors that potentially affects the agreement on grading normal clivus since it initiates and modulates the marrow conversion differently with respect to the age spectrum. Moreover, radiologist's experiences also potentially affect the reading of grading the clivus. Although studies about the influence of radiologists' experiences on reading MRI of clivus have not been studied, previous research emphasizes the importance of the level of experience on reading images [8–11]. It is worth noting that sharing experiences and finding the consensus on the grading of clivus among radiologists will gain more accuracy in reading; therefore, the overall agreement might be improved.

To our knowledge, there is limited information about the studies of agreement among radiologists on reading MRI calvarial marrow especially clivus. Our study might be the first to evaluate the agreement among radiologists on reading clivus on MRI. The representativeness of the sample under study is one of the strengths of our study since we randomly sampled individuals who have been performed brain MRI in Radiology department from a large cohort of 1717 individuals during January 2015 to October 2019. In addition, all radiologists had received a set of images as a training procedure before commencing the study. However, some considerations should be noted. Firstly, we chose the established grading criteria from Kimura et al. and Oyar et al. which employed subjective assessment and the exact low SI portion of clivus area was unknown [2,6]. Although visual assessment can cause impair agreement, our study confirmed a good agreement among radiologists based on this grading system. Secondly, we only evaluated the patients with presumed normal bone marrow. Pathologic processes such as tumor invasion or marrow reconversion often can cause changes in the cellular composition of the marrow. Thus, our study cannot be generalized to patients with pathologic conditions. Further study on the agreement of grading clivus in these patients is recommended. Thirdly, our study mainly focused on the marrow of clivus. One previous study revealed that marrow conversion in segments of cranial bone were different; for instance, fatty conversion of the presphenoid marrow occurred earlier than in the other parts of the skull base [12]. In midline sagittal T1-weighted image, some parts of calvarium and upper cervical spine were also visualized that might be useful information. Future study in other parts of calvarium is suggested.

5. Conclusions

The agreement of grading the clivus among radiologists were good. The visual grading criteria used in our setting is sufficient to distinguish the marrow conversion of clivus. The consensus reading should be made whenever clivus Grade II is read. Further efforts to derive system for evaluating Grade II remain warranted.

Ethics approval and consent to participate

This retrospective study involving human participants was in accordance with the ethical standards of Naresuan university research ethics committee, Thailand.

Funding source

This article was financially supported by Faculty of Medicine, Naresuan University (grant number MD2563C005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Funding statement

We certify that there was no source of funding for any of the author involved in preparation of this manuscript.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

We gratefully acknowledge the hard work, efficiency, and devotion of our imaging technicians, which made this work possible.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejro.2022.100395](https://doi.org/10.1016/j.ejro.2022.100395).

References

- [1] T. Laor, D. Jaramillo, MR imaging insights into skeletal maturation: What is normal? *Radiology* 250 (2009) 28–38, <https://doi.org/10.1148/radiol.2501071322>.
- [2] F. Kimura, K.W. Kim, H. Friedman, E.J. Russell, R. Breit, MR imaging of the normal and abnormal clivus, *Am. J. Roentgenol.* 155 (1990) 1285–1291, <https://doi.org/10.2214/ajr.155.6.2122682>.
- [3] Y. Okada, S. Aoki, A.J. Barkovich, K. Nishimura, D. Norman, B.O. Kjos, R.C. Brasch, Cranial bone marrow in children: assessment of normal development with MR imaging, *Radiology* 171 (1989) 161–164, <https://doi.org/10.1148/radiology.171.1.2928520>.
- [4] A. Bayramoğlu, Ü. Aydingöz, M. Hayran, H. Öztürk, M. Cumhuri, Comparison of qualitative and quantitative analyses of age-related changes in clivus bone marrow on MR imaging, *Clin. Anat.* 16 (2003) 304–308, <https://doi.org/10.1002/ca.10065>.
- [5] E. Olcu, M. Arslan, V. Sabanciogullar, I. Salk, Magnetic resonance imaging of the clivus and its age-related changes in the bone marrow, *Iran. J. Radiol.* 8 (2012) 156–161, <https://doi.org/10.5812/iranjrad.4494>.
- [6] O. Oyar, F. Gövsa, R.N. Sener, G. Kayalioglu, Assessment of normal clivus related to age with magnetic resonance imaging, *Surg. Radiol. Anat.* 18 (1996) 47–49, <https://doi.org/10.1007/BF03207762>.
- [7] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, *Biometrics.* 33 (1977) 159–174.
- [8] F.H.J. Elsholtz, S.-R. Ro, S. Shnayien, C. Erxleben, H.-C. Bauknecht, J. Lenk, L.-A. Schaafs, B. Hamm, S.M. Niehues, Inter- and Intra-reader Agreement of NI-RADS in the Interpretation of Surveillance Contrast-Enhanced CT after Treatment of Oral Cavity and Oropharyngeal Squamous Cell Carcinoma, *Am. J. Neuroradiol.* 41 (2020) 859 LP–859865, <https://doi.org/10.3174/ajnr.A6529>.
- [9] F.O. Henes, M. Groth, T.A. Bley, M. Regier, J.V. Nüchtern, H. Itrich, A. Tresszl, G. Adam, P. Bannas, Quantitative assessment of bone marrow attenuation values at MDCT: an objective tool for the detection of bone bruise related to occult sacral insufficiency fractures, *Eur. Radiol.* 22 (2012) 2229–2236, <https://doi.org/10.1007/s00330-012-2472-8>.
- [10] H. Geijer, M. Geijer, Added value of double reading in diagnostic radiology, a systematic review, *Insights Imaging* 9 (2018) 287–301, <https://doi.org/10.1007/s13244-018-0599-0>.
- [11] M. Kaup, J.L. Wichmann, J.-E. Scholtz, M. Beeres, W. Kromen, M.H. Albrecht, T. Lehnert, M. Boettcher, T.J. Vogl, R.W. Bauer, Dual-energy CT-based display of bone marrow edema in osteoporotic vertebral compression fractures: impact on diagnostic accuracy of radiologists with varying levels of experience in correlation to MR imaging, *Radiology* 280 (2016) 510–519, <https://doi.org/10.1148/radiol.2016150472>.
- [12] A. Taccone, M. Oddone, M. Occhi, A. Dell'Acqua, M.A. Ciccone, MRI “road-map” of normal age-related bone marrow - I. Cranial bone and spine, *Pediatr. Radiol.* 25 (1995) 588–595, <https://doi.org/10.1007/BF02011825>.