

A Quantitative MRI Index for Assessing the Severity of Hippocampal Sclerosis in Temporal Lobe Epilepsy

Wanchen Dou

Peking Union Medical College Hospital

Lei Zhao

BrainNow Research Institute <https://orcid.org/0000-0001-5125-974X>

Changbao Su

Peking Union Medical College Hospital

Qiang Lu

Peking Union Medical College Hospital

Qi Liu

Peking Union Medical College Hospital

Jinzhu Guo

Peking Union Medical College Hospital

Yuming Zhao

Peking Union Medical College Hospital

Yishan Luo

BrainNow Research Institute

Lin Shi

Chinese University of Hong Kong

Yiwei Zhang

Peking Union Medical College Hospital

Renzhi Wang (✉ wangrz@126.com)

Peking Union Medical College Hospital

Feng Feng

Peking Union Medical College Hospital

Research article

Keywords: temporal lobe epilepsy, hippocampal sclerosis, grading scale, quantitative MRI, index

Posted Date: November 11th, 2019

DOI: <https://doi.org/10.21203/rs.2.17148/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Imaging on April 25th, 2020.

See the published version at <https://doi.org/10.1186/s12880-020-00440-z>.

Abstract

Background Hippocampal sclerosis (HS) is associated with post-surgery outcome in patients with temporal lobe epilepsy (TLE), and an automated method that quantifies HS severity is still lacking. Here, we aim to propose an MRI-based HS index (HSI) that integrates hippocampal volume and FLAIR signal to measure the severity of HS.

Methods Forty-two pre-surgery TLE patients were included retrospectively, with T1-weighted (T1W) and FLAIR images acquired from each subject. Two experienced neurosurgeons (W.D. and C.S.) and one neurologist (Q.L.) rated HS severity, where both the hippocampal volume loss and increased FLAIR signal were considered to provide a four-class grading scale (normal, mild, moderate and severe). A consensus of HS severity for each subject was made by voting among the three visual rating results. Regarding the automatic quantification, the hippocampal volume was quantified by AccuBrain on T1W image, and the FLAIR signal of hippocampus was calculated as the mean intensity of hippocampal region on the FLAIR image (normalized by the mean intensity of gray matter). To fit the HSI from visual rating, we applied ordinal regression with the voted visual rating as the dependent variable, and hippocampal volume and FLAIR signal as the independent variables. The HSI was calculated by weighting the predictive probabilities of the four-class grading scales from ordinal regression.

Results The intra-class correlation coefficient (single measure) of the three raters was 0.806. The generated HSI was significantly correlated with the visual rating scales (W.D.: 0.823, Q.L.: 0.817, C.S.: 0.717). HSI scores differentiated well between each pair of groups defined by the agreed HS visual rating (normal vs. mild: $p < 0.001$, mild vs. moderate: $p < 0.001$, moderate vs. severe: $p = 0.001$).

Conclusions The proposed HSI was consistent with visual rating scales from neurosurgeons and sensitive to HS severity. This MRI-based index may help to evaluate HS severity in clinical practice. Further validations are needed to associate HSI with post-surgery outcomes.

Background

Hippocampal sclerosis (HS) is the most common pathology underlying medically intractable temporal lobe epilepsy (TLE) [1]. The best option to enable the TLE patients seizure-free is surgical resection, especially when HS is present [2]. Moreover, HS severity was also found to associate with the long-term post-surgery outcome of TLE [3]. In general, HS is confirmed with histopathological specimens from the resected tissue, where an international consensus of HS classification (ILAE) has also been built in the recent years [4]. However, the neuropathological examinations are invasive, and the presence of HS can only be identified after the surgical resection is done with pathological approach. To this end, noninvasive quantitative MRI has played an important role in TLE diagnosis and treatment of epilepsy since the early 1990s [5], particularly when surgery is being considered [6].

The typical MRI features of HS include hippocampal volume loss on T1-weighted (T1W) imaging and increased signal intensity on T2-weighted (or FLAIR) imaging [6, 7]. As hippocampal atrophy can also be

an age-related change and increased T2 signal also presents with lesions other than HS [8], these two features achieved better detection of HS when used in combination [9–11]. However, unlike neuropathological examinations that report HS pattern or severity (e.g. ILAE type [4] or Wyler grading score[1]), MRI-based HS evaluation was generally binary (presence of HS or not). This may result from the lack of visual rating-based grading scale on MRI, which can serve as the target of model fitting with quantitative MRI features (as predictors). Although high resolution MRI [12–14] (e.g. 7T MRI) studies have associated hippocampal substructure MRI features with pathological grading scales, it cannot be applied in practice due to its limited access in clinical routine.

In this regard, we proposed a study to quantify HS severity based on MRI features with two steps. First, a four-class grading scale of HS based on MRI was constructed, where two experienced neurosurgeons (with over 20 years of epileptic surgery) and one experienced neurologist at the Epileptic Center were involved for the visual ratings. Second, with the agreed HS rating on a TLE cohort, we fitted an ordinal regression model to map the association between the MRI features (hippocampal atrophy and increased T2 signal) and the HS grading score, resulting in a continuous index of HS. As the reliability of hippocampal segmentation tool also matters in detection of HS [15], we applied the recently validated software AccuBrain[16] for the quantification of hippocampal volume and T2 signal. The generated index of HS severity was compared with the visual rating categories of the raters for validation.

Methods

Subjects

We recruited 42 patients (21 male, 12–52 years, mean \pm SD age 29 ± 10 years) with TLE that was refractory to medical therapy as diagnosed at the Epilepsy Center of Peking Union Medical College Hospital. The data collection was conducted retrospectively. The inclusion criteria were detailed as follows: (1) the semiology of the patients matched typical clinical manifestation of temporal lobe epilepsy; (2) the patients had undergone brain MRI (both 3D T1-weighted (T1W) scan and 2D FLAIR scan) before surgery; (3) the patients had evidence of ictal or interictal epileptiform activity arising from temporal lobe monitored with video electroencephalography (VEEG); (4) the patients had pre-surgery PET/CT examination that indicated hypometabolism in unilateral or bilateral temporal lobe. (5) Indications of surgery for TLE were made through multi-department consultations for the patients. The exclusion criteria were: (1) patients with extra-temporal lobe epilepsy; (2) presence of other lesions on MRI, such as cavernous hemangioma and ganglioglioma; (3) the patients who only had 3D T1W scan or 2D FLAIR scan.

Data acquisition

A 3T MR imaging system (Discovery MR 750 scanner, GE Medical Systems) was used for image acquisition. Three-dimensional sagittal T1W images were acquired using a gradient echo sequence (BRAVO) with following parameters: repetition time (TR) = 7.404 ms, echo time (TE) = 2.82 ms, flip angle

(FA) = 12°, inversion time (TI) = 400 ms, matrix = 512×512, FOV = 512×512 mm², number of slices = 180, slice thickness = 1.0 mm, no gap, spatial resolution = 0.4688×0.4688×1 mm³. Fluid-attenuated inversion recovery (FLAIR) sequences were acquired at the coronal plane with TR = 12037.5 ms, TE = 123 ms, FA = 111°, TI = 2200 ms, matrix = 512×512, number of slices = 38, and slice thickness = 4.0 mm, spatial resolution = 0.4297×0.4297×4 mm³.

Visual rating

Visual rating was performed by two neurosurgeons (W. D. and C. S.) with over 20 years' experience in TLE surgery and an experienced neurologist (Q. L.) at the Epilepsy Center. A four-class grading system (0, normal; 1, mild; 2, moderate; 3 severe) was applied according to the raters' experiences in surgery and each rater rated the all the TLE patients independently. For each rater, the patients were rated twice at different time points, and a final judgement was confirmed based on these two ratings.

Classic signs of MRI-based HS visual rating include reduction of hippocampal volume on T1W images and increased signal observed in FLAIR images, which are generally examined on coronal slices perpendicular to the long axis of the hippocampus [6, 17]. The severity of HS (grading score) was further estimated as follows: (1) on T1W image, the patient will be rated as having severe HS if the hippocampal volume decreases by over 50%, and rated as moderate or mild HS for a reduction of hippocampal volume by 25%~50% or <25% respectively; if no obvious volume reduction is found, the subject will be rated as normal as far as T1W is concerned; (2) On FLAIR image, the intensity within hippocampus was compared with the surrounding grey matter to rate the strength of increased signal, as part of HS severity from the perspective of FLAIR-based rating; (3) the ratings of T1W and FLAIR images were finally combined to a synthetic four-class grading score (empirically defined weights for a linear combination: 0.8 for T1W and 0.2 for FLAIR).

Images were also carefully examined by the investigators in light of the clinical and EEG data of each patient [10]. When the ratings from the three raters were ready, a simple voting was performed to make an agreed HS grading for each individual. To evaluate the inter-rater reliability of the three raters, the intraclass correlation coefficient (ICC) was also tested using Cronbach's alpha [18].

Image processing

Hippocampal volumetric analysis were performed with AccuBrain[®] (on T1W MRI scans), which has achieved the best performance in hippocampal segmentation among the existing automatic brain segmentation tools in a recent validation study[16]. The absolute volume of hippocampus was normalized by intracranial volume (ICV), and the resulting hippocampal volume ratio (% of ICV) was used as one predictor for model construction in the following section. Grey matter (GM) tissue of the whole brain was also segmented on T1W with AccuBrain. The masks of hippocampus (both left and right) and GM from segmentation were projected from the T1W image to the FLAIR image of the subject with affine

registration [19]. The resulting masks were used to outline the hippocampal and GM regions, and thus to calculate the average intensity of hippocampus and GM on FLAIR image. The relative intensity of hippocampus on FLAIR (i.e. the ratio of mean intensity of hippocampus to the mean intensity of GM) was used as another predictor for model construction.

Statistical analysis

The T1W-based predictor (hippocampal volume) and the FLAIR-based predictor (relative intensity of hippocampus) were first used to fit a model with the grading scale from visual rating as the outcome. Here, we flipped the right hippocampus to the left to enlarge the data size for model fitting (i.e. 84 hippocampal data in 42 TLE patients). As the grading scale is an ordinal variable, we applied ordinal regression (OR) to for model fitting. Among the OR approaches in literature [20], we selected the threshold models (where an unobserved continuous variable is assumed to underlie the ordinal response) for examination, including Proportional Odds Model [21] (POM, a linear model extended from logistic regression), Neural Network based POM [21] (NNPOM, nonlinear generalization of POM), Support Vector Ordinal Regression with Explicit Constraints [22] (SVOREX), SVOR with Implicit Constraints [22] (SVORIM), Kernel Discriminant Learning for Ordinal Regression [23] (KDLOR), and Reduction from ordinal regression to binary support vector machine [24] (REDSVM). The implementation of these methods is available for free from the website of the authors of the OR review paper [20] (<http://www.uco.es/grupos/ayrna/orreview>).

A 5-fold cross-validation (where each fold was once used for testing and the remaining folds for training) was performed when fitting each of the five models. The model performance was comprehensively evaluated with three metrics: (1) mean accuracy (ACC), a rigorous metric that indicates the ratio of correctly predicted cases (of any class) among all cases; (2) mean absolute error (MAE)[25], the average deviation between predicted and actual targets in number of categories; (3) Spearman's rank correlation (R), nonparametric correlation between the predicted and actual categories. The best model is expected to have the optimal balanced prediction performance (ACC+R-MAE) after parameter optimization. Also, the optimal model should also have as few parameters to train as possible to control overfitting[26] given the small sample size in this study.

To further generalize the ordinal outcome of the optimal model to a continuous HS index (HSI), we applied a weighted linear combination of the probability of each category (as generated by the optimal OR model):

$$\text{HSI} = (P_1 + 2*P_2 + 3*P_3)/3, \text{ (Eq. 1)}$$

Where P_1 , P_2 and P_3 indicate the probability of mild, moderate and severe HS for a specific subject; the probability of no HS (normal, P_0) is not included for HSI calculation as it would have a term of $(0*P_0)$ that has no contribution to the formula. The HSI ranges from 0 to 1, where a larger HSI indicates more severe HS.

To evaluate the performance of HSI in differentiating visual rating-based HS categories, we compared the HSI of different visual rating-based categories with Mann-Whitney U test (especially between normal vs. mild, mild vs. moderate and moderate vs. severe HS). Also, Spearman’s rank correlation tests were performed to evaluate the consistency among HSI and the rating scores of the three raters.

Results

According to the voted visual rating, there were 4 subjects with no HS (age 30.5 ± 11.0 years), 11 subjects with bilateral HS (age 26.6 ± 11.6 years), 15 subjects with left HS (age 27.3 ± 9.0 years), and 11 subjects with right HS (age 33.6 ± 8.2 years) in our TLE cohort. By flipping the hippocampal data of the study cohort, each hippocampus was rated with the four-class grading scale independently, with the group size and relevant characteristics (e.g. age or onset years of TLE corresponding to a specific hippocampi) shown in **Table 1**. Regarding the voting of the visual ratings, at least two raters had agreement on HS for all the hippocampus, and the raters had better agreement on normal HS (21 of 35 normal HS agreed by all the three raters) than the other categories (e.g. 3 of 16 mild HS agreed by all the three raters). As shown in **Table 2**, the three raters achieved comparable intra-rater reproducibility (mean ICC of single measures 0.832) and the inter-rater reliability was relatively high (0.806 for single measures of ICC).

Table 1. Characteristics of the TLE patients in different HS categories

| | Normal | Mild HS | Moderate HS | Severe HS |
|-------------------------|----------------|-----------------|----------------|-----------------|
| Group size by voting | 35 | 19 | 22 | 8 |
| Agreed by 2 raters | 14 | 16 | 18 | 7 |
| Agreed by 3 raters | 21 | 3 | 4 | 1 |
| Age, y, mean \pm SD | 30.1 ± 9.2 | 29.4 ± 11.0 | 27.9 ± 9.2 | 27.8 ± 12.2 |
| Gender, male | 18 | 7 | 13 | 4 |
| Onset, y, mean \pm SD | 13.3 ± 8.2 | 11.4 ± 7.4 | 14.1 ± 9.8 | 15.1 ± 7.5 |

The displayed characteristics correspond to the flipped hippocampal data. HS, hippocampal sclerosis.

Table 2. Intra-rater reproducibility and inter-rater reliability of the three raters

| | ICC single measures (95% CI) | ICC average measures (95% CI) |
|-----------------------------|------------------------------|-------------------------------|
| Intra-rater reproducibility | | |
| Rater W.D | 0.874 (0.825-0.912) | 0.954 (0.934-0.969) |
| Rater Q.L. | 0.878 (0.831-0.915) | 0.956 (0.937-0.930) |
| Rater C.S. | 0.744 (0.657-0.816) | 0.897 (0.852-0.930) |
| Inter-rater reliability | 0.806 (0.736-0.863) | 0.926 (0.893-0.950) |

The displayed ICC coefficients were all significant at the level of $p < 0.001$. ICC, intraclass correlation coefficient; CI, confidence interval.

Model selection with cross-validations

In general, the performances of the various ordinal regression models were very similar (Table 3). Although the KDLOR model achieved the best performance (highest in terms of ACC+R-MAE), it requires three parameters for model training. In this regard, we selected the traditional POM, which achieved similar performance with KDLOR and did not need any model parameters to train.

Table 3. Performance of different ordinal regression models in 5-fold cross-validations

| Model | ACC | MAE | R | ACC+R-MAE | Optimized parameter(s) |
|--------|--------|--------|--------|-----------|--------------------------------|
| POM | 0.6919 | 0.3199 | 0.8428 | 1.2148 | --- |
| NNPOM | 0.6801 | 0.3551 | 0.8391 | 1.1641 | H=5 |
| SVOREX | 0.6794 | 0.3441 | 0.8472 | 1.1825 | k=10, c=10 |
| SVORIM | 0.6912 | 0.3324 | 0.8310 | 1.1898 | k=10, c=10 |
| KDLOR | 0.7037 | 0.3081 | 0.8527 | 1.2483 | k=1, c=0.1, u=10 ⁻⁶ |
| REDSVM | 0.7037 | 0.3199 | 0.8386 | 1.2224 | k=10, c=10 |

The mean ACC, MAE and R during the 5-fold cross-validations were displayed for each optimized ordinal regression (OR) model. The searching ranges of model parameters follow the OR review paper[20]: $H \in \{5, 10, 20, 30, 40\}$, $k \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$, $c \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$, $u \in \{10^{-6}, 10^{-5}, \dots, 10^{-2}\}$. ACC, accuracy; MAE, mean absolute error; R, Spearman's rank correlation; POM, proportional odds model; NNPOM, Neural network based on POM; SVOREX, support vector ordinal regression with explicit constraints; SVORIM, support vector ordinal regression with implicit constraints; KDLOR, kernel discriminant learning for ordinal regression; REDSVM, reduction from ordinal regression to binary support vector machine. H, the number of hidden neurons; k, the width of Gaussian kernel function; c, cost parameter of all SVM methods; u, additional parameter of KDLOR that is intended to avoid singularities in the covariance matrices[20].

Comparison of HSI with visual rating

As POM was selected based on the performance of 5-fold cross-validation, we fitted a POM model based on the entire database and used the generated probability of each category to calculate the HSI according to Equation (1). As shown in Figure 1, the fitted HSI well differentiated the four categories, with little overlap of boxplots between any of the adjacent two categories. As confirmed by the Mann-Whitney U tests, the HSI was significantly larger in a more severe HS category between its adjacent HS category (NC < mild HS, $p < 0.001$; mild HS < moderate HS, $p < 0.001$; moderate HS < severe HS, $p = 0.001$). In addition, the fitted HSI had high correlations with the grading scales of each rater (mean 0.786), which were comparable to the inter-rater correlations of the grading scales (mean 0.806) as shown in Table 4.

Table 4. Correlation between HSI and the HS categories identified from different raters

| Spearman's rank correlation | HSI | Rater D.W. | Rater Q.L. | Rater C.S. |
|-----------------------------|-------|------------|------------|------------|
| HSI | 1.000 | 0.823 | 0.817 | 0.717 |
| Rater D.W. | 0.823 | 1.000 | 0.878 | 0.777 |
| Rater Q.L. | 0.817 | 0.878 | 1.000 | 0.762 |
| Rater C.S. | 0.717 | 0.777 | 0.762 | 1.000 |

The correlation coefficients shown above were all significant at the level of $p < 0.001$.

Discussion

In this study, for the first time we constructed an automated MRI-based index (HSI) to depict the HS severity in medically intractable TLE patients. The HSI was generated by fitting an optimized OR model with hippocampal volume and hippocampal FLAIR signal as the predictors and the proposed four-class grading scales from visual inspection as the reference outcome.

Hippocampal volume and T2 or FLAIR signal have been widely applied (either individually [27, 28] or in combination [7, 9–11]) to differentiate hippocampus with HS from that without HS in TLE patients, but no study has applied these MRI features to evaluate HS severity, which is associated with postoperative outcome in TLE patients [3]. This might result from the lack of golden standard for visual inspection of HS severity as a reference, since routine brain MRI generally provides very limited information compared with the neuropathological scans where visual grading scales are available [1, 4]. Although high resolution MRI scans (e.g. 7T MRI) have been reported to depict pathological information of HS [12–14], they can rarely be used in clinical practice due to their high cost and poor accessibility. In this regard, we turned to integrate the experiences of visual inspection for HS (on 3T MRI) from neurosurgeon and neurologist for a four-class grading scale to rate HS severity. Although the proposed grading scale was relatively subjective, where the extent of hippocampal atrophy and FLAIR hyperintensity in hippocampus was scored only based on the experiences of the raters, we found that at least two of the three raters agreed on the rating of all the hippocampi of the cohort with a high inter-rater reliability (ICC: 0.814 for single measures and 0.929 for average measures).

Based on the voted rating scale from the three raters, we tried a series of OR models [20] to map the raters' subjective but consistent experiences of visual inspection for HS into a quantitative model, with hippocampal volume and FLAIR intensity of hippocampus (% of GM intensity) as the independent variables. A comprehensive model selection was made among almost all the currently available threshold models of OR [20], with 5-fold cross-validations to increase the generalizability of the result. Although KDLOR achieved the best performance, it also required the most parameters to train, which induced more chances of overfitting [26]. Finally, we selected the POM model which ranked 3rd in model performance and did not fall behind much compared with the top two models (KDLOR and REDSVM, *Table 3*), as it required no model parameters to train. In POM, the ordinal outcome is generated by weighting the probability of each category, which provides more information than the ordinal outcome itself. In this regard, we integrated these probabilities into a continuous HS score (i.e. HSI, as shown in Eq. 1), where a

more severe HS category would contribute more to the HSI score assuming that the probability of each category was the same. The resulting HSI well differentiated the visual rating-based HS categories (*Figure 1, Table 4*), which indicated that the automated MRI features-driven index indirectly learned the experiences of neurosurgeon and neurologist on HS rating.

There are several limitations to this study that should be considered. First, the sample size of the study cohort is relatively small, and an external validation dataset is not available. However, we applied cross-validations and comprehensive model selections with least parameters to train, aiming to make the results as generalizable as possible. Further validation should be made in a larger cohort to test the reliability of the proposed POM model to calculate HSI. Of note, age and gender were not included in model construction for the HSI, as they were not correlated with the grading score in the study cohort. Future work might also include these covariates if they present correlation with the target HS severity when a larger sample for model construction is available. Second, the proposed grading scale of HS severity from visual inspection was relatively subjective. However, the involved three raters achieved good intra-rater reproducibility and inter-rater reliability (*Table 2*), which indicated that a consistent visual rating can be realized under this grading scheme among experienced neurosurgeon and neurologist at Epilepsy Center. In addition, regarding the type of T2 signal (one of MRI features to measure HS), we applied normalized FLAIR intensity instead of T2 relaxometry, which was more widely used in the existing studies, due to the lack of this MRI sequence of our study cohort. As a recent study reported that T2 relaxometry was more sensitive to normalized FLAIR intensity to detect HS (when hippocampal atrophy was not considered) [29], future work should aim to test if T2 relaxometry would also contribute to a better evaluation of HS severity when hippocampal atrophy is also considered in the model. Furthermore, a comprehensive pathological validation of the proposed HSI was not available in this study, as only a small subset of the participants had the surgical samples that could be used for evaluation of HS pathology (*Supplementary Table S1*), which primarily resulted from the retrospective design of this study. Finally, post-surgery outcomes of the TLE cohort were not yet available, and our following validations will focus on the associations between HSI and post-surgery outcomes to evaluate the use of HSI in clinical practice.

Conclusions

In summary, this study proposed an MRI-based index (HSI) based on automatic quantification of hippocampal atrophy and increased FLAIR intensity to measure the severity of HS, which showed high consistency with visual rating scales from experienced neurosurgeons and neurologist. The HSI may help to evaluate HS severity in clinical practice, which warrants further validations in a larger sample of TLE patients with accessible post-surgery outcomes.

List Of Abbreviations

ACC: accuracy

FA: Flip angle

FLAIR: Fluid-attenuated inversion recovery

FOV: Field of view

GM: Grey matter

HS: Hippocampal sclerosis

HSI: Hippocampal sclerosis index

ICC: Intraclass correlation coefficient

ICV: Intracranial volume

ILAE: International League Against Epilepsy

KDLOR: Kernel discriminant learning for ordinal regression

MAE: mean absolute error

NC: Normal control

NNPOM: Neural network based proportional odds model

OR: Ordinal regression

POM: Proportional odds model

REDSVM: Reduction from ordinal regression to binary support vector machine

SVOREX: Support vector ordinal regression with explicit constraints

SVORIM: Support vector ordinal regression with implicit constraints

T1W: T1-weighted

TE: Echo time

TI: Inversion time

TLE: Temporal lobe epilepsy

TR: Repetition time

Declarations

Ethics approval and consent to participate

The present study was approved by the Institutional Ethics Committee of Peking Union Medical College Hospital. Written informed consent was obtained from all individual participants included in the study.

Consent for publication

All patients agreed to participate in this study and provided written informed consent.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

L. S. is the director of BrainNow Medical Technology Limited. L. Z. and Y. L. are employees of BrainNow Medical Technology Limited. All other authors report no financial relationships with commercial interests.

Funding

This study was financially supported by 2016-I2M-1-004 from the CAMS Innovation Fund for Medical Sciences (CIFMS). The funding bodies had no role in the design, data collection, data analysis or data interpretation and no role in writing the manuscript.

Authors' contributions

Study design: WD, LS, RW and FF; Literature research: LZ, YZ and LS; Data collection: JG and YZ; Visual rating and pathological analysis: WD, CS and QL; Data analysis: LZ and YL; manuscript preparation: WD, LZ and LS. All authors have read and approved the manuscript.

Acknowledgments

Not applicable.

References

1. Watson C, Nielsen SL, Cobb C, Burgerman R, Williamson B. Pathological grading system for hippocampal sclerosis: correlation with magnetic resonance imaging-based volume measurements of the hippocampus. *J Epilepsy*. 1996;9(1):56–64.

2. Wieshmann UC, Larkin D, Varma T, Eldridge P. Predictors of outcome after temporal lobectomy for refractory temporal lobe epilepsy. *Acta Neurol Scand*. 2008;118(5):306–12.
3. Na M, Ge H, Shi C, Shen H, Wang Y, Pu S, et al. Long-term seizure outcome for international consensus classification of hippocampal sclerosis: a survival analysis. *Seizure*. 2015;25:141–6.
4. Blumcke I, Thom M, Aronica E, Armstrong DD, Bartolomei F, Bernasconi A, et al. International consensus classification of hippocampal sclerosis in temporal lobe epilepsy: a Task Force report from the ILAE Commission on Diagnostic Methods. *Epilepsia*. 2013;54(7):1315–29.
5. Jack CR, Jr., Bentley MD, Twomey CK, Zinsmeister AR. MR imaging-based volume measurements of the hippocampal formation and anterior temporal lobe: validation studies. *Radiology*. 1990;176(1):205–9.
6. Bernasconi A, Cendes F, Theodore WH, Gill RS, Koepp MJ, Hogan RE, et al. Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: A consensus report from the International League Against Epilepsy Neuroimaging Task Force. *Epilepsia*. 2019.
7. Van Paesschen W, Sisodiya S, Connelly A, Duncan JS, Free SL, Raymond AA, et al. Quantitative hippocampal MRI and intractable temporal lobe epilepsy. *Neurology*. 1995;45(12):2233–40.
8. Labate A, Gambardella A, Aguglia U, Condino F, Ventura P, Lanza P, et al. Temporal lobe abnormalities on brain MRI in healthy volunteers: a prospective case-control study. *Neurology*. 2010;74(7):553–7.
9. Van Paesschen W, Connelly A, King MD, Jackson GD, Duncan JS. The spectrum of hippocampal sclerosis: a quantitative magnetic resonance imaging study. *Ann Neurol*. 1997;41(1):41–51.
10. Coan AC, Kubota B, Bergo FP, Campos BM, Cendes F. 3T MRI quantification of hippocampal volume and signal in mesial temporal lobe epilepsy improves detection of hippocampal sclerosis. *AJNR Am J Neuroradiol*. 2014;35(1):77–83.
11. Winston GP, Vos SB, Burdett JL, Cardoso MJ, Ourselin S, Duncan JS. Automated T2 relaxometry of the hippocampus for temporal lobe epilepsy. *Epilepsia*. 2017;58(9):1645–52.
12. Coras R, Milesi G, Zucca I, Mastropietro A, Scotti A, Figini M, et al. 7T MRI features in control human hippocampus and hippocampal sclerosis: an ex vivo study with histologic correlations. *Epilepsia*. 2014;55(12):2003–16.
13. Urbach H, Huppertz HJ, Schwarzwald R, Becker AJ, Wagner J, Bahri MD, et al. Is the type and extent of hippocampal sclerosis measurable on high-resolution MRI? *Neuroradiology*. 2014;56(9):731–5.
14. Mueller SG, Laxer KD, Barakos J, Cheong I, Garcia P, Weiner MW. Subfield atrophy pattern in temporal lobe epilepsy with and without mesial sclerosis detected by high-resolution MRI at 4 Tesla: preliminary results. *Epilepsia*. 2009;50(6):1474–83.

15. Pardoe HR, Pell GS, Abbott DF, Jackson GD. Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia*. 2009;50(12):2586–92.
16. Abrigo J, Shi L, Luo Y, Chen Q, Chu WCW, Mok VCT, et al. Standardization of hippocampus volumetry using automated brain structure volumetry tool for an initial Alzheimer's disease imaging biomarker. *Acta Radiol*. 2019;60(6):769–76.
17. Malmgren K, Thom M. Hippocampal sclerosis—origins and imaging. *Epilepsia*. 2012;53 Suppl 4(s4):19–33.
18. Cronbach LJ. Coefficient alpha and the internal structure of tests. *psychometrika*. 1951;16(3):297–334.
19. Jafari-Khouzani K, Elisevich K, Patel S, Smith B, Soltanian-Zadeh H. FLAIR signal and texture analysis for lateralizing mesial temporal lobe epilepsy. *Neuroimage*. 2010;49(2):1559–71.
20. Gutierrez PA, Perez-Ortiz M, Sanchez-Monedero J, Fernandez-Navarro F, Hervas-Martinez C. Ordinal Regression Methods: Survey and Experimental Study. *IEEE Trans Knowl Data Eng*. 2016;28(1):127–46.
21. McCullagh P. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1980;42(2):109–27.
22. Chu W, Keerthi SS. Support vector ordinal regression. *Neural Comput*. 2007;19(3):792–815.
23. Sun B-Y, Li J, Wu DD, Zhang X-M, Li W-B. Kernel Discriminant Learning for Ordinal Regression. *IEEE Trans Knowl Data Eng*. 2010;22(6):906–10.
24. Lin HT, Li L. Reduction from cost-sensitive ordinal ranking to weighted binary classification. *Neural Comput*. 2012;24(5):1329–67.
25. Cruz-Ramírez M, Hervás-Martínez C, Sánchez-Monedero J, Gutiérrez PA. Metrics to guide a multi-objective evolutionary algorithm for ordinal classification. *Neurocomputing*. 2014;135:21–31.
26. Domingos P. A few useful things to know about machine learning. *Communications of the ACM*. 2012;55(10):78.
27. Watson C, Cendes F, Fuerst D, Dubeau F, Williamson B, Evans A, et al. Specificity of volumetric magnetic resonance imaging in detecting hippocampal sclerosis. *Arch Neurol*. 1997;54(1):67–73.
28. Mueller SG, Laxer KD, Schuff N, Weiner MW. Voxel-based T2 relaxation rate measurements in temporal lobe epilepsy (TLE) with and without mesial temporal sclerosis. *Epilepsia*. 2007;48(2):220–8.
29. Rodionov R, Bartlett PA, He C, Vos SB, Focke NK, Ourselin SG, et al. T2 mapping outperforms normalised FLAIR in identifying hippocampal sclerosis. *Neuroimage Clin*. 2015;7:788–91.

Figures

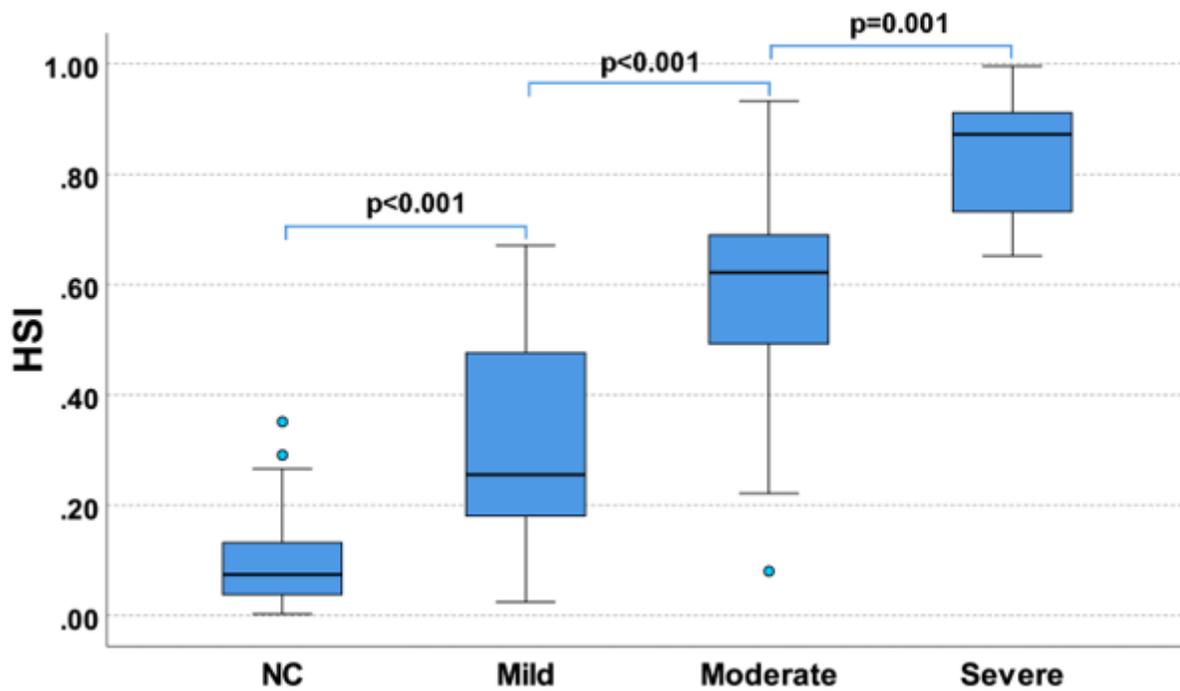


Figure 1

Boxplot of HSI among visual rating-based categories

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.docx](#)