Original Article

Imaging signs for preoperative diagnosis of dural ossification in patients with thoracic ossification of the ligamentum flavum: a blind, randomized diagnostic accuracy study

Guanghui Chen[1,2,3#^], Liyuan Tao[4#^], Zhongqiang Chen[1,2,3], Weishi Li[1,2,3^], Chunli Song[1,2,3^], Woquan Zhong[1,2,3], Yu Jiang[1,2,3], Xinhu Guo[1,2,3], Tianqi Fan[1,2,3], Shuai Jiang[1,2,3], Chuiguo Sun[1,2,3^]

[1]Department of Orthopedics, Peking University Third Hospital, Beijing, China; [2]Beijing Key Laboratory of Spinal Disease Research, Beijing, China; [3]Engineering Research Center of Bone and Joint Precision Medicine, Ministry of Education, Beijing, China; [4]Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China

Contributions: (I) Conception and design: G Chen, L Tao, C Sun; (II) Administrative support: C Sun, Z Chen, W Li, C Song; (III) Provision of study materials or patients: W Zhong, Y Jiang, X Guo, C Sun; (IV) Collection and assembly of data: G Chen, T Fan, S Jiang; (V) Data analysis and interpretation: G Chen, L Tao, C Sun; (VI) Manuscript writing: All authors; (VII) Final approval of manuscripts: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Chuiguo Sun, MD. Department of Orthopedics, Peking University Third Hospital, 49 North Garden Road, Haidian District, Beijing 100191, China; Beijing Key Laboratory of Spinal Disease Research, Beijing, China; Engineering Research Center of Bone and Joint Precision Medicine, Ministry of Education, Beijing, China. Email: sunchuiguo@163.com.

Background: Dural ossification (DO) is the leading cause of surgery-related dural tear in patients with ossification of the ligamentum flavum (OLF). An accurate preoperative diagnosis of DO is conducive to the selection of appropriate surgical methods. Although several imaging signs, such as Banner cloud sign (BCs), tram-track sign (TTs), and comma sign (Cs) have been proposed for the preoperative diagnosis of DO, their diagnostic value has not been well studied. The aim of this study was to explore the diagnostic value of BCs, TTs, and Cs, and provide evidence-based data for their clinical application.

Methods: This is a blind, randomized diagnostic study using retrospectively collected data from 102 consecutive patients who were diagnosed with OLF and underwent decompression surgery between January 2018 and June 2019. A total of 8 surgeons with different qualifications were recruited to read these imaging signs to identify the presence of DO. Surgical records were used as the reference standard. Sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve (AUC) were used to evaluate the diagnostic accuracy of each imaging sign and their different combinations.

Results: Of the 102 patients, 21 were diagnosed with DO. BCs had a significantly higher diagnostic accuracy than TTs and Cs, with the AUC of 0.704, 0.607, and 0.593, respectively. The specificity of BCs, Cs, TTs, and their combination in diagnosing DO was 91.5%, 92.1%, 68.3%, and 62.2%, respectively. In the combined diagnostic test, the results showed that the combined diagnosis accuracy of BCs and Cs was the highest, and the AUC was 0.738. The combination of BCs, Cs, and TTs increased the sensitivity of diagnosing DO (77.5%), but did not improve the diagnostic accuracy, and the AUC was 0.699.

Conclusions: BCs had higher diagnostic accuracy than TTs and Cs. BCs and Cs were highly specific for DO, whereas TTs could be confusing due to their non-specific presentations. The combination of BCs, TTs, and Cs improved the sensitivity of DO diagnosis, but not the specificity and accuracy.

^ ORCID: Guanghui Chen, 0000-0001-6031-0180; Liyuan Tao, 0000-0003-3497-1326; Weishi Li, 0000-0001-9512-5436; Chunli Song, 0000-0002-3690-9457; Chuiguo Sun, 0000-0003-3478-8361.

Keywords: Imaging signs; ossification of the ligamentum flavum (OLF); dural ossification (DO); diagnostic accuracy

Submitted May 08, 2023. Accepted for publication Nov 10, 2023. Published online Nov 30, 2023.
doi: 10.21037/qims-23-634
View this article at: https://dx.doi.org/10.21037/qims-23-634

Introduction

Ossification of the ligamentum flavum (OLF) is the leading cause of thoracic spinal stenosis (TSS), accounting for approximately 72% of patients requiring thoracic decompression surgery (1,2). Its natural course is progressive and it responds poorly to conservative treatment, making surgery the only effective method. However, the high incidence of surgery-related complications, such as cerebrospinal fluid (CSF) leakage and neurological deterioration, has always been a major concern for most surgeons (3).

Dural ossification (DO), characterized by ossified dura mater fused with ossified ligamentum flavum, is the main cause of CSF leakage (4). It is reported that 78.8–85.7% of patients with postoperative CSF leakage have DO (5,6). In addition, some other complications, such as pseudocyst, nervous system infection, and wound dehiscence secondary to CSF leakage, can also lead to catastrophic events if not managed properly (7). Therefore, accurate preoperative diagnosis of DO and adequate preparation to prevent intraoperative dural tear are crucial.

Currently, methods for preoperative diagnosis of DO are limited because of the rarity of the disease. Although several imaging signs, such as tram-track sign (TTs) (8), comma sign (Cs) (8), Bridge sign (Bs) (9), and magnetic resonance imaging (MRI)-T2 ring sign (10), have been proposed, their diagnostic accuracy has not been well studied (11) and their diagnostic value has not been fully explored.

To improve the preoperative diagnostic accuracy of DO, we have previously introduced a new imaging sign, namely, Banner cloud sign (BCs) (6), and published a protocol to investigate its diagnostic value (12). Its clinical characteristics and correlation with DO have also been preliminary studied (6). However, its diagnostic value has not yet been fully investigated. In light of this, we conducted this blind, randomized diagnostic accuracy study to explore and compare the diagnostic values of BCs, TTs, and Cs, aiming to provide evidence-based data for their clinical application. We present this article in accordance with the STARD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-23-634/rc).

Methods

Study population

The present study was approved by the Ethics Committee of Peking University Third Hospital (IRB00006761-M2019494) and conducted according to the principles of the Declaration of Helsinki (as revised in 2013). Written informed consent was provided by all participants. The protocol has been published and registered on the Chinese Clinical Trials website (ChiCTR2000030380) (12).

In this study, a consecutive of 130 patients diagnosed with TSS who underwent decompression surgery in our center between January 2018 and June 2019 were preliminary analyzed, and 102 patients who met the inclusion and exclusion criteria were finally enrolled. The inclusion criteria were as follows: (I) patients with thoracic OLF undergoing posterior decompression surgery; (II) operation was conducted between January 2018 and June 2019; and (III) complete medical records and operation notes. The exclusion criteria were as follows: (I) patients who had ossification of the posterior longitudinal ligament (OPLL), diffuse idiopathic skeletal hyperostosis (DISH), skeletal fluorosis, and degenerative TSS; (II) patients who underwent revision surgery or had missing image data; and (III) patients with thoracic trauma, infection, tumor, or deformity.

Study design

The study was carried out according to our protocol (12). Before the study, 8 surgeons with different qualifications (2 attending surgeons, 2 clinical fellows, 2 residents, and 2 interns), representing different experiences in reading images, were trained for 3 sessions to familiarize themselves with the typical imaging features of BCs, TTs, and Cs

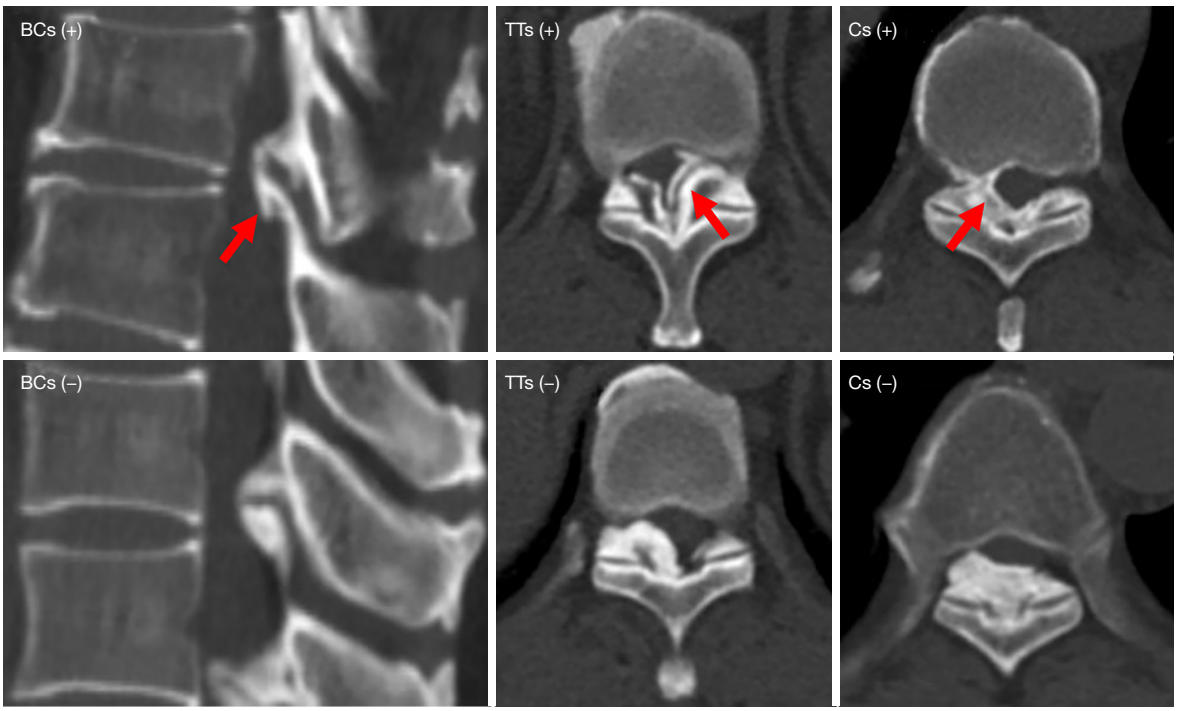

Figure 1 The representative CT images of BCs (left panel), TTs (middle panel) and Cs (right panel). For each typical imaging sign, (+) indicates positive and (–) indicates negative. The red arrows indicate the differences between these imaging signs. BCs, Banner cloud sign; TTs, tram-track sign; Cs, comma sign; CT, computed tomography.

based on preoperative reconstructed computed tomography (CT). The typical characteristics of these signs are shown in *Figure 1*. The other 2 surgeons who did not participate in the experiment were responsible for the data collection and anonymization. Then, the 102 data sets, including 428 segments of OLF, were randomly numbered by statistical experts (TLY) and distributed to each observer for reading. Each observer could read only 1 imaging sign at a time to determine the presence of DO, so each observer would read the data sets 3 times. The presence of DO was determined according to description of Kaufman *et al.* (13), and surgical records were used as a reference standard. Sensitivity and specificity data were presumed to be based on correlation of imaging findings to the presence or absence of DO based on review of surgical records. Patient enrollment and the study protocol are shown in *Figure 2*.

Outcome assessment

The primary outcome was to evaluate and compare the diagnostic accuracy of each imaging sign by sensitivity, specificity, Youden index, and the area under the receiver operating characteristic (ROC) curve (AUC). The secondary outcomes were to assess the time and confidence required for each observer to diagnose DO based on different imaging signs, so as to determine the universality and ease of mastery of each imaging sign. The diagnostic accuracy of different combinations of imaging signs was also calculated and compared.

Statistical analysis

In this study, online data management software REDCap (Vanderbilt University, Nashville, TN, USA) was used for data collection and management. The software SPSS 26.0 (IBM Corp., Armonk, NY, USA) was used for statistical calculation. The quantitative data conforming to the normal distribution were expressed as mean ± standard deviation, and the counting data were described by the number of cases (%). Multiple group comparisons were performed using a one-way analysis of variance (ANOVA) followed by *post-hoc* test using the Bonferroni method. The comparison of counting data was conducted by chi-square test. Sensitivity, specificity, Yuden index, ROC curve, and AUC were used

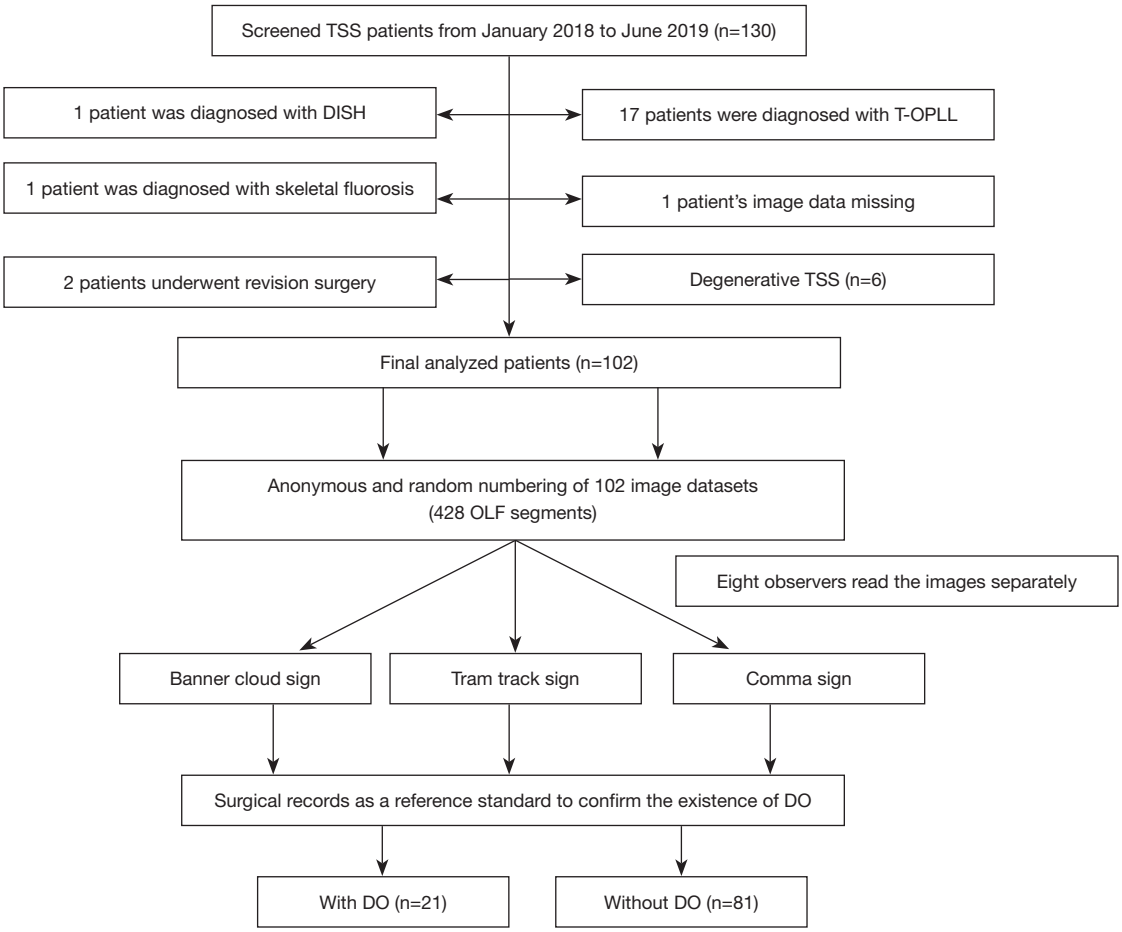

Figure 2 Flowchart of patient enrollment and study protocol. TSS, thoracic spinal stenosis; DISH, diffuse idiopathic skeletal hyperostosis; T-OPLL, thoracic ossification of the posterior longitudinal ligament; OLF, ossification of the ligamentum flavum; DO, dural ossification.

to compare the diagnostic accuracy of different signs, and a combined test was used to compare the diagnostic accuracy of different combinations of imaging signs. Interclass correlation coefficients (ICC) were calculated to assess interobserver reliability. Multivariable logistic regression analysis and decision tree analysis [growing method: chi-square automatic interaction detector (CHAID)] were used to explore the diagnostic effect of different imaging signs, and R Software ("rms" package) was used to construct the nomogram risk assessment tool and calibration curve. A P value <0.05 was considered statistically significant.

Results

Patient characteristics

In this study, 102 patients (428 segments of OLF) were enrolled and analyzed. The cohort consisted of 57 males and 45 females with an average age of 55±11 years. A total of 21 patients were diagnosed with DO, and the incidence of DO in OLF patients was 20.6% (21/102). Of the OLF segments, 46 were accompanied by DO, and the incidence was 10.8% (46/428). In segment distribution, 44.9% (192/428) of OLF and 60.9% (28/46) of DO were located in the lower thoracic spine (T9–12). The detailed patient demographic data are shown in *Table 1*.

Diagnostic accuracy assessment

The diagnostic accuracy of different imaging signs is shown in *Table 2*. BCs showed the highest diagnostic accuracy, with AUC of 0.704 [95% confidence interval (CI): 0.671 to 0.738], followed by TTs and Cs with AUC of 0.607 (95% CI: 0.575 to 0.639) and 0.593 (95% CI: 0.559 to 0.627),

Table 1 Clinical characteristics of all OLF patients

Items	Value
Age (years)	55±11
Sex, M/F	57/45
Segment distribution of OLF	
Upper thoracic spine (T1–4)	26.2% (112/428)
Middle thoracic spine (T5–8)	29.0% (124/428)
Lower thoracic spine (T9–T12)	44.9% (192/428)
Segment distribution of DO	
Upper thoracic spine (T1–4)	13.0% (6/46)
Middle thoracic spine (T5–8)	26.1% (12/46)
Lower thoracic spine (T9–T12)	60.9% (28/46)

Data are presented as mean ± standard deviation, n, or % (n/N). OLF, ossification of the ligamentum flavum; M, male; F, female; DO, dural ossification.

respectively. The AUC of BCs was significantly larger than that of TTs and Cs (P<0.01), but there was no significant difference between TTs and Cs (P=0.356). Although the overall sensitivity of these imaging signs was low, the diagnostic specificity of BCs and Cs for DO was relatively high at 91.5% and 92.1%, respectively. In addition, there was no statistically significant difference in AUC when observers reading each imaging sign, and the P values of BCs, TTs and Cs were 0.682, 0.709, and 0.716, respectively.

Confidence, time, and consistency analysis

There were significant differences in the confidence of different observers in reading different imaging signs. Briefly, observers showed the highest confidence in identifying Cs, followed by BCs and TTs (P<0.01). Interns had the highest confidence in evaluating BCs and Cs, and residents had the lowest confidence (P<0.001). Fellows

Table 2 Comparison of diagnostic accuracy between different imaging signs and observers

Items	Sensitivity (%)	Specificity (%)	Youden index (%)	AUC (95% CI)	P value
BCs	49.4	91.5	40.9	0.704 (0.671–0.738)	<0.01*
Interns	38.2	94.0	32.2	0.661 (0.592–0.730)	
Residents	46.2	93.7	39.9	0.699 (0.623–0.766)	0.682†
Fellows	65.1	83.0	48.1	0.740 (0.679–0.802)	
Attendings	48.9	94.9	43.8	0.719 (0.652–0.786)	
TTs	53.1	68.3	21.4	0.607 (0.575–0.639)	0.356‡
Interns	31.5	81.8	13.3	0.566 (0.500–0.633)	
Residents	59.3	50.1	9.4	0.547 (0.485–0.609)	0.709†
Fellows	69.8	64.3	34.1	0.670 (0.611–0.730)	
Attendings	52.2	77.9	30.1	0.651 (0.586–0.715)	
Cs	26.5	92.1	18.6	0.593 (0.559–0.627)	
Interns	21.3	96.8	18.1	0.591 (0.522–0.660)	
Residents	41.8	86.0	27.8	0.639 (0.572–0.705)	0.716†
Fellows	27.9	88.9	16.8	0.584 (0.515–0.653)	
Attendings	14.4	96.7	11.1	0.556 (0.489–0.623)	

*, indicates there was significant difference between BCs and TTs/Cs; †, indicates there was no significant between different observers in reading each sign. ‡, indicates there was no significant difference between TTs and Cs. AUC, area under the curve; CI, confidence interval; BCs; Banner cloud sign; TTs, tram-track sign; Cs, comma sign.

Table 3 Comparison of confidence and time taken of different observers in reading different imaging signs

Items	Summary		Subgroup analysis				
	Mean	P value	Interns	Residents	Fellows	Attendings	P value
Confidence*							
BCs	8.47±1.00	ref	9.01±0.83	7.94±1.01	8.45±0.77	8.50±1.05	<0.001
TTs	7.73±1.31	<0.001	8.21±0.87	6.40±1.02	8.65±0.94	7.69±1.17	<0.001
Cs	8.53±1.34	0.008	9.58±0.63	7.21±1.15	8.65±0.85	8.71±1.34	<0.001
Time taken† (s)							
BCs	32.18±24.27	ref	34.97±27.96	31.08±27.16	33.37±19.88	29.40±20.53	<0.001
TTs	30.82±24.70	0.002	31.79±21.82	27.75±16.67	33.61±31.50	30.22±26.25	<0.001
Cs	22.52±17.82	<0.001	20.11±13.70	24.00±16.33	24.60±22.83	21.42±16.96	<0.001

Data are presented as mean ± standard deviation. *, indicates the confidence of observers in reading imaging signs of a single segment; †, indicates the time taken by the observers in reading imaging signs of a single segment. BCs, Banner cloud sign; TTs, tram-track sign; Cs, comma sign.

showed higher confidence in identifying TTs, followed by interns, attendings, and residents (P<0.001). In addition, there were significant differences in time required to obtain the results between observers of different qualifications and when using different imaging signs (P<0.001). The detailed information is shown in *Table 3*.

ICC analysis showed that the interobserver reliability of BCs was almost in perfect agreement, whereas that of TTs and Cs was in substantial agreement, and the ICC values were 0.81 (95% CI: 0.78 to 0.84), 0.79 (95% CI: 0.75 to 0.82), and 0.77 (95% CI: 0.74 to 0.81), respectively. The ICC values of different seniority observers were also different when observing the same imaging sign. The highest (0.68; 95% CI: 0.61 to 0.74) and lowest (0.26; 95% CI: 0.10 to 0.39) ICC were present in the identification of TTs, indicating a significant inconsistency in the readings of this imaging sign (*Figure 3*).

Diagnostic accuracy of different combinations of imaging signs

The diagnostic accuracy of different combinations is shown in *Table 4* and *Figure 4*. The highest diagnostic accuracy was obtained when combining BCs with Cs; the sensitivity, specificity, and AUC were 62.1%, 85.5%, and 0.738 (95% CI: 0.707 to 0.769), respectively. However, the combination of BCs with TTs showed no significant improvement in diagnostic specificity and AUC compared to BCs. Similarly, the combination of all these three imaging signs achieved an AUC of 0.699 (95% CI: 0.671 to 0.726), which was not higher than that of the combination of BCs and Cs. These results suggest that the combined use of TTs improved the diagnostic sensitivity of DO, but not the diagnostic accuracy.

Clinical scenarios

For clinical application, a nomogram model was constructed based on the multivariable logistic regression analysis. The results showed that BCs had the greatest influence on diagnostic accuracy, followed by Cs and TTs (*Figure 5*). The value of each of these imaging signs was given a score on the point scale axis. A total score could be calculated by adding each single score and projecting the total score to the lower total point scale. The calibration curve indicated that the predicted risk of DO was roughly consistent with the actual risk (*Figure 6*). Decision tree analysis showed similar results to those of logistic regression analysis, and highlighted the diagnostic effect of BCs in DO diagnosis. The diagnostic accuracy of the decision tree model was 77.5% for positive patients and 62.2% for negative patients. The detailed diagnostic accuracy for each step is shown in *Figure 7*.

Discussion

DO represents a difficult condition in the surgical management of OLF (14,15). Its incidence in OLF patients varies from 11% to 62% and it is the major cause of dura tear and CSF leakage (6,16). Previous studies (5,6,11) reported that DO causes about 78.8–85.7% of CSF leakage

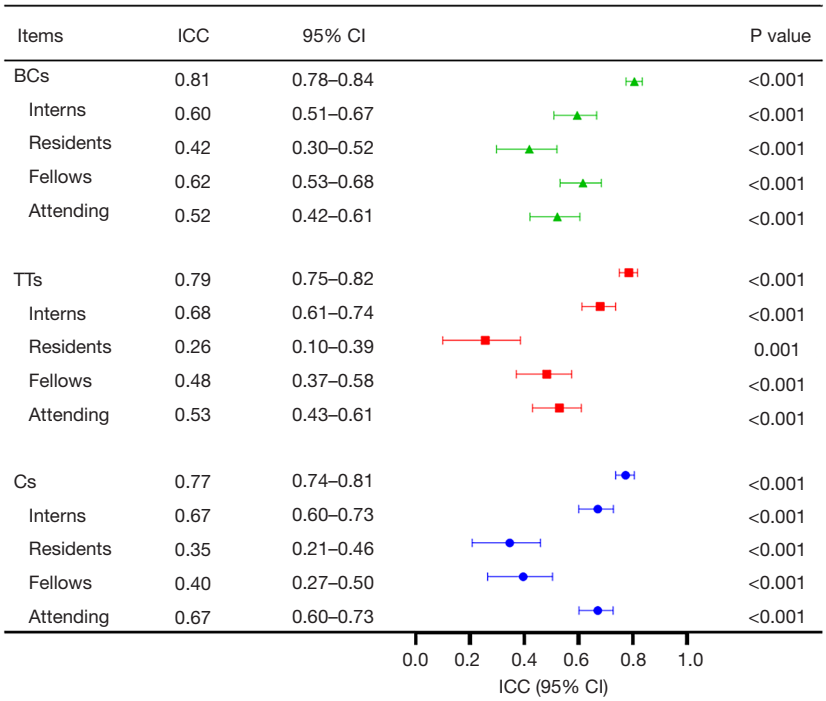

Figure 3 Comparison of ICC between different observers in reading each imaging sign. ICC, interclass correlation coefficients; BCs, Banner cloud sign; TTs, tram-track sign; Cs, comma sign; CI, confidence interval.

Table 4 Diagnostic accuracy of different combinations of imaging signs

Combinations	Sensitivity (%)	Specificity (%)	Youden index (%)	AUC (95% CI)
BCs or Cs	62.10	85.50	47.60	0.738 (0.707–0.769)
BCs or TTs	72.50	65.30	37.80	0.689 (0.660–0.717)
Cs or TTs	61.50	64.90	26.40	0.632 (0.601–0.663)
BCs, Cs or TTs	77.50	62.20	39.70	0.699 (0.671–0.726)

AUC, area under the curve; CI, confidence interval; BCs, Banner clouds sign; Cs, comma sign; TTs, tram-track sign.

in OLF patients. In this study, we found that 20.6% of OLF patients had at least one segment of DO. The results were consistent with previous studies (6,11,17), suggesting that patients with OLF have a higher incidence of DO, and accurate preoperative diagnosis of DO was very important.

Several methods for diagnosing DO have been described, which can be divided into two categories according to their research methods: one is based on spinal canal-occupying ratio (SCOR), and the other is based on imaging signs. Using SCOR or Sato classification to predict the possibility of DO may be applied to most cases because the more spinal canal is occupied, the more severe dura mater compression is, and the greater is the risk of developing DO (12-14,18). However, this kind of method cannot give a "yes" or "no" diagnostic opinion for a certain case but merely offer a possibility and, therefore, cannot meet the needs of clinicians for preoperative surgical planning.

In terms of the imaging signs, TTs and Cs, introduced by Muthukumar *et al.* in 2009 (8), are widely used due to their simplicity and convenience. However, as there were only 9 cases in their study, the diagnostic accuracy of these imaging signs has not been studied. In a later retrospective study, Sun *et al.* (5) reported that the diagnostic specificity of TTs was only 59%. Furthermore, in 2016, Li *et al.* (9) reported four types of false TTs, which confused the surgeon's diagnosis and led to low diagnostic accuracy. Yet, so far,

no high-quality studies have been conducted to explore its diagnostic value.

In the previous study, we introduced a computed tomography (CT)-based imaging sign, BCs, and preliminarily analyzed its clinical characteristics and diagnostic accuracy (6). However, its diagnostic value has also not been fully studied. Therefore, we conducted this blind, randomized, large-sample study to explore the diagnosis values of these imaging signs and to provide evidence-based data to guide their clinical application.

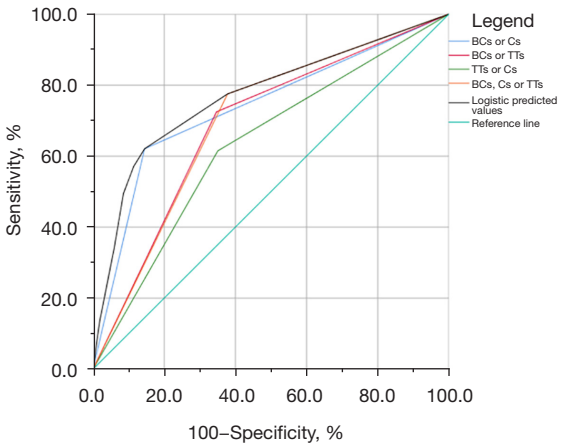

Figure 4 ROC curves showed the diagnostic accuracy of different combinations of imaging signs. BCs, Banner cloud sign; TTs, tram-track sign; Cs, comma sign; ROC, receiver operating characteristic.

In terms of diagnostic accuracy, BCs was much higher than TTs and Cs (70.4%, 60.7%, and 59.3%, respectively). A common feature of these imaging signs was that their diagnostic sensitivity was relatively low; nearly 50% of DO were misdiagnosed by BCs or TTs and 74.5% by Cs. The reason for the low sensitivity of Cs was that only a few cases had typical Cs. These results suggest that there must be some other atypical imaging findings that have not been detected. Further studies are needed to explore, discover, and perfect the manifestations of these signs. Although the sensitivity of these signs was low, the specificity of BCs and Cs was relatively high (91.5% and 92.1%, respectively), indicating that if BCs or Cs are observed in CT images, the existence of DO should be highly suspected. In contrast, the specificity of TTs was low (68.3%), suggesting that this imaging sign needs further improvement. In addition, since there was no difference in diagnostic accuracy among different observers, we concluded that these signs could be

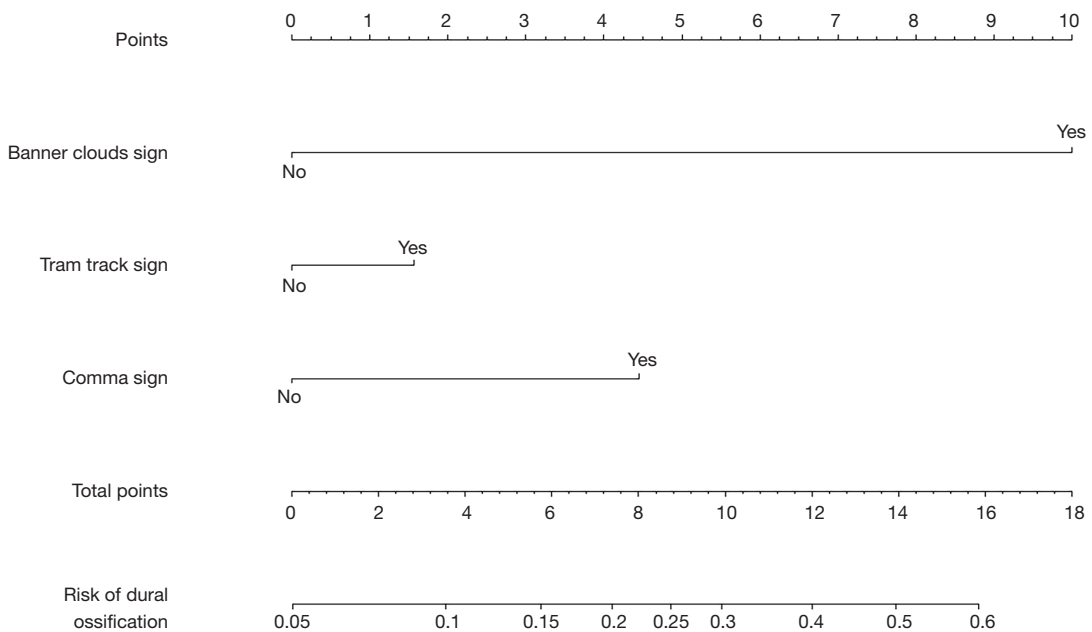

Figure 5 A nomogram model predicting the risk of DO in patients with OLF. The value of each of these imaging signs was given a score on the point scale axis. The total score was calculated by adding each single score and projecting the total score to the lower total point scale. DO, dural ossification; OLF, ossification of the ligamentum flavum.

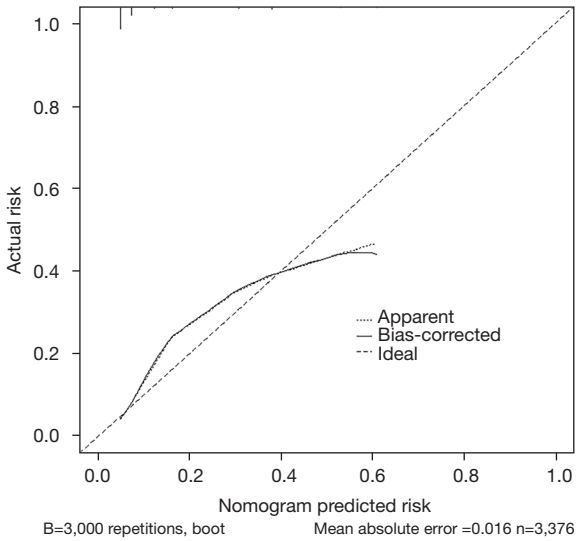

Figure 6 The calibration curves for the nomogram. The x-axis represents the nomogram-predicted probability and y-axis represents the actual probability of DO. Perfect predication would correspond to the 45° dashed line. The dotted line represents all the number of OLF segments read by 8 observers (n=3,376), and the solid line is bias-corrected by bootstrapping (B=3,000 repetitions), indicating observed nomogram performance. DO, dural ossification; OLF, ossification of the ligamentum flavum.

easily mastered after a short training session, regardless of the surgeon's film reading experience.

Since an ideal diagnostic method should be universal and accessible to clinicians at different levels of experience, we compared the confidence and time taken by each observer to identify these signs. We found that surgeons had the highest confidence in identifying Cs and the lowest confidence in identifying TTs. These results suggest that Cs were easier to master because of their typical imaging findings, whereas the atypical nature of TTs confuses surgeons. We also found that surgeons took longer to identify BCs than the other two imaging signs. Altogether, these findings suggest that surgeons more easily mastered Cs, followed by BCs and then TTs. Moreover, to evaluate the inter-observer reliability of different methods, the ICC analysis was performed. The results showed that surgeons have a relatively high degree of consistency in identifying these imaging signs. However, the highest and lowest ICC appeared simultaneously in identifying TTs, indicating that surgeons have a very different understanding of TTs and that TTs were not easy to master.

Previous studies have shown that combining multiple diagnostic methods can improve the diagnostic accuracy (9,17). Li *et al.* (9) reported that combining TTs, Cs, and

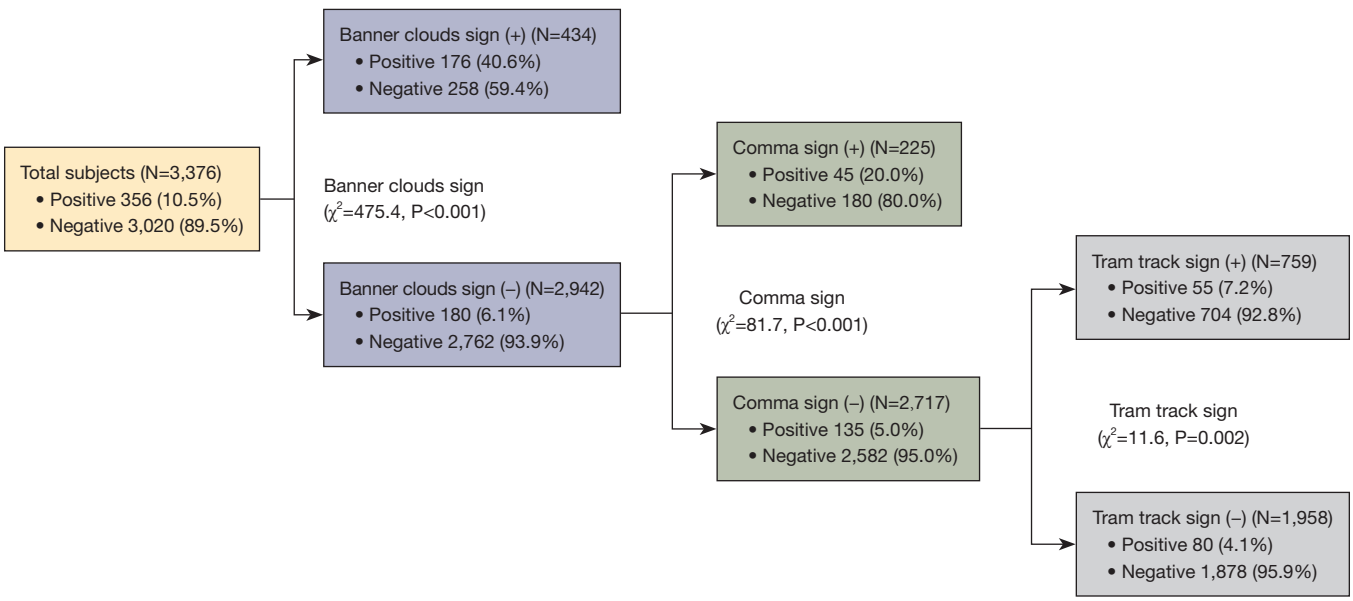

Figure 7 Decision tress model presents the optimal diagnostic strategy and the detailed diagnostic accuracy of each imaging sign in each step.

Bs could achieve a high diagnostic accuracy of 94.21%. Zhou *et al.* (17) constructed a thoracic OLF (TOLF)-DO grading system using unilateral spinal canal occupying rate and imaging signs (TTs and Cs) and reported a diagnostic accuracy of 87.3%. In this study, we found that compared with BCs alone, the combined use of these three imaging signs improved the sensitivity (77.5%) significantly but not the specificity (62.2%) and accuracy (AUC, 0.699).

The lower sensitivity indicates that DO may still exist despite the absence of one or more of these signs, and surgeons should keep in mind the possibility that DO is present in order to respond to an emergency. Nevertheless, the results are encouraging, because we have a 70% probability of correctly diagnosing DO through the combined application of these three methods. In clinical practice, high diagnostic sensitivity is necessary, because it can minimize the rate of missed diagnosis of DO as much as possible, thereby reminding surgeons to make sufficient preoperative preparation and surgical planning, and reducing the occurrence of postoperative complications. In the future, some new imaging signs or methods should be developed to improve the sensitivity of DO diagnosis.

For clinical application, logistic regression and decision tree analysis were performed, and the results showed that BCs had the highest diagnostic value, followed by Cs and TTs. By using the constructed nomogram model, the occurrence of DO can be accurately predicted; therefore, we speculate that this model might be a useful tool for DO diagnosis. According to the decision tree model, the diagnosis of DO should be considered firstly according to the presence of BCs, followed by TTs and then Cs.

To the best of our knowledge, this is the first blind, randomized, large-sample study that comprehensively explored the diagnostic values of BCs, TTs, and Cs. The sample size was large, and the study was carried out strictly in accordance with our protocol. The results have provided reliable evidence-based data to guide their application in DO diagnosis. In addition, the nomogram and decision tree model provide a reasonable and feasible solution for the clinical diagnosis of DO. However, several limitations in this study should be addressed. Since TTs and Cs are currently the most commonly used imaging signs to diagnose DO, we only investigated the diagnostic values of our proposed imaging sign, BCs, and TTs; Cs, and some other signs such as Bs and MRI-T2 ring signs should be included in future studies. In addition, the diagnostic value of the nomogram and decision tree models in predicting DO needs further clinical validation. In future studies, other more typical imaging sings or new diagnostic methods should be explored and developed to improve the diagnostic accuracy of DO.

Conclusions

BCs have higher diagnostic accuracy than TTs and Cs. BCs and Cs are highly specific and easy to master, whereas TTs can be confusing for the diagnosis of DO due to their non-specific presentations. The combined use of BCs, TTs, and Cs can significantly improve the diagnostic sensitivity of DO, but not for the specificity and diagnostic accuracy. The nomogram and decision tree model may be useful tools for diagnosing DO. Further prospective, large-scale studies are needed to validate our findings, and some other more typical imaging features or diagnostic methods, such as artificial intelligence, should be explored and developed to improve the diagnosis accuracy of DO.

Acknowledgments

We thank Baoliang Zhang and Xiaoxi Yang of Peking University Third Hospital for their assistance in data collection.

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 82072479, 81772381, 81874031), Peking University Third Hospital Talent Incubation Fund (No. Y77477-03), and China Postdoctoral Science Foundation (No. 2022M720293).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-23-634/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-634/coif). The authors have no conflict of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The present study was approved by the Ethics Committee of Peking University Third Hospital and conducted according to the principles of the Declaration of Helsinki (as revised in

2013). The protocol has been registered on the Clinical Trials website (ChiCTR2000030380). Written informed consent was provided by all participants.

References

1. Hou X, Sun C, Liu X, Liu Z, Qi Q, Guo Z, Li W, Zeng Y, Chen Z. Clinical Features of Thoracic Spinal Stenosis-associated Myelopathy: A Retrospective Analysis of 427 Cases. Clin Spine Surg 2016;29:86-9.
2. Chen G, Fan T, Yang X, Sun C, Fan D, Chen Z. The prevalence and clinical characteristics of thoracic spinal stenosis: a systematic review. Eur Spine J 2020;29:2164-72.
3. Hou X, Chen Z, Sun C, Zhang G, Wu S, Liu Z. A systematic review of complications in thoracic spine surgery for ossification of ligamentum flavum. Spinal Cord 2018;56:301-7.
4. Hu P, Yu M, Liu X, Liu Z, Jiang L, Wei F, Chen Z. Cerebrospinal Fluid Leakage after Surgeries on the Thoracic Spine: A Review of 362 Cases. Asian Spine J 2016;10:472-9.
5. Sun XZ, Chen ZQ, Qi Q, Guo ZQ, Sun CG, Li WS, Zeng Y. Diagnosis and treatment of ossification of the ligamentum flavum associated with dural ossification: clinical article. J Neurosurg Spine 2011;15:386-92.
6. Chen G, Chen Z, Li W, Jiang Y, Guo X, Zhang B, Tao L, Song C, Sun C. Banner cloud sign: a novel method for the diagnosis of dural ossification in patients with thoracic ossification of the ligamentum flavum. Eur Spine J 2022;31:1719-27.
7. Sun X, Sun C, Liu X, Liu Z, Qi Q, Guo Z, Leng H, Chen Z. The frequency and treatment of dural tears and cerebrospinal fluid leakage in 266 patients with thoracic myelopathy caused by ossification of the ligamentum flavum. Spine (Phila Pa 1976) 2012;37:E702-7.
8. Muthukumar N. Dural ossification in ossification of the ligamentum flavum: a preliminary report. Spine (Phila Pa 1976) 2009;34:2654-61.
9. Li B, Qiu G, Guo S, Li W, Li Y, Peng H, Wang C, Zhao Y. Dural ossification associated with ossification of ligamentum flavum in the thoracic spine: a retrospective analysis. BMJ Open 2016;6:e013887.
10. Prasad GL. Thoracic spine ossified ligamentum flavum: single-surgeon experience of fifteen cases and a new MRI finding for preoperative diagnosis of dural ossification. Br J Neurosurg 2020;34:638-46.
11. Yu L, Li B, Yu Y, Li W, Qiu G, Zhao Y. The Relationship Between Dural Ossification and Spinal Stenosis in Thoracic Ossification of the Ligamentum Flavum. J Bone Joint Surg Am 2019;101:606-12.
12. Chen G, Zhang B, Tao L, Chen Z, Sun C. The diagnostic accuracy of CT-based "Banner cloud sign" for dural ossification in patients with thoracic ossification of the ligamentum flavum: a prospective, blinded, diagnostic accuracy study protocol. Ann Transl Med 2020;8:1606.
13. Kaufman AB, Dunsmore RH. Clinicopathological considerations in spinal meningeal calcification and ossification. Neurology 1971;21:1243-8.
14. Osman NS, Cheung ZB, Hussain AK, Phan K, Arvind V, Vig KS, Vargas L, Kim JS, Cho SK. Outcomes and Complications Following Laminectomy Alone for Thoracic Myelopathy due to Ossified Ligamentum Flavum: A Systematic Review and Meta-Analysis. Spine (Phila Pa 1976) 2018;43:E842-8.
15. Miyakoshi N, Shimada Y, Suzuki T, Hongo M, Kasukawa Y, Okada K, Itoi E. Factors related to long-term outcome after decompressive surgery for ossification of the ligamentum flavum of the thoracic spine. J Neurosurg 2003;99:251-6.
16. Aizawa T, Sato T, Sasaki H, Kusakabe T, Morozumi N, Kokubun S. Thoracic myelopathy caused by ossification of the ligamentum flavum: clinical features and surgical results in the Japanese population. J Neurosurg Spine 2006;5:514-9.
17. Zhou SY, Yuan B, Chen XS, Li XB, Zhu W, Jia LS. Imaging grading system for the diagnosis of dural ossification based on 102 segments of TOLF CT bone-window data. Sci Rep 2017;7:2983.
18. Zhao Y, Xiang Q, Jiang S, Lin J, Wang L, Sun C, Li W. Incidence and risk factors of dural ossification in patients with thoracic ossification of the ligamentum flavum. J Neurosurg Spine 2023;38:131-8.

Cite this article as: Chen G, Tao L, Chen Z, Li W, Song C, Zhong W, Jiang Y, Guo X, Fan T, Jiang S, Sun C. Imaging signs for preoperative diagnosis of dural ossification in patients with thoracic ossification of the ligamentum flavum: a blind, randomized diagnostic accuracy study. Quant Imaging Med Surg 2024;14(2):1466-1476. doi: 10.21037/qims-23-634

