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

Guanghui Chen[1,2#^], Baoliang Zhang[1,2#], Liyuan Tao[3], Zhongqiang Chen[1,2], Chuiguo Sun[1,2]

[1]Department of Orthopedics, Peking University Third Hospital, Beijing, China; [2]Beijing Key Laboratory of Spinal Disease Research, Beijing, China; [3]Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China

[#]These authors contributed equally to this work.

Correspondence to: Chuiguo Sun. Department of Orthopedics, Peking University Third Hospital, Peking University, #49 North Garden Road, Haidian District, Beijing 100191, China. Email: sunchuiguo@163.com.

Background: Ossification of the ligamentum flavum (OLF) is the most common cause of thoracic spinal stenosis, which responds poorly to conservative treatment. Thus, surgery is the only effective treatment for OLF. The existence of dural ossification (DO) makes surgery challenging and increases the risk of intra-/ post-operative complications. To date, several methods have been proposed to identify DO, but either the diagnostic accuracy is low or the feasibility is poor. Therefore, the aim of this study was to propose a new imaging sign (Banner cloud sign, BCs), evaluate the accuracy of BCs in the diagnosis of DO, and provide reliable evidence-based data for its application in clinical practice.

Methods: A prospective, blinded, diagnostic accuracy study will be conducted to assess and compare the accuracy of BCs in the diagnosis of DO with other radiological signs [Tram track sign (TTs) and Comma sign (Cs)]. A total of 120 patients diagnosed with OLF who underwent decompression at the Peking University Third Hospital between January 2018 and June 2019 will be enrolled. Patients' medical records and imaging data will be retrieved from the hospital database server. An observational group consisting of six spine surgeons (with different seniority levels) and two epidemiological researchers will read the patients' images to identify typical imaging signs and determine the presence of DO. Surgical records will be reviewed to confirm the presence of DO, and the results will serve as the reference standard for estimating accuracy. The primary outcome of the study is to determine the accuracy of BCs for DO diagnosis, and the secondary outcome is to compare the sensitivity, specificity, area under the receiver operating characteristic (ROC) curve and inter-observer reliability of each imaging sign. The time taken and level of confidence in DO diagnosis of each observer will also be compared.

Discussion: This study represents the first large-scale investigation of the diagnostic value of BCs, TTs and Cs in the diagnosis of DO, and will provide convincing evidence about their clinical application.

Trial registration: Registered on 29 February 2020. Trial number is ChiCTR2000030380.

Keywords: Ossification of the ligamentum flavum (OLF); dural ossification (DO); complication; diagnosis; tram track sign

^ ORCID: 0000-0001-6031-0180.

Page 2 of 8

Chen et al. CT-based "Banner cloud sign" for DO

Introduction

Ossification of the ligamentum flavum (OLF) is the most common cause of thoracic spinal stenosis, especially in East Asian countries with a prevalence rate of 3.8–63.9% (1,2). It has progressive natural course and responds poorly to conservative treatment. Thus surgery is the only effective treatment for OLF. Previous studies showed that decompressive surgery on the thoracic spine was very challenging and the existence of dural ossification (DO) increased the surgical difficulty (3-5). Complications secondary to DO, such as neurological deterioration and cerebrospinal fluid (CSF) leakage, were frequently reported.

Our previous study showed that DO was the leading cause of CSF leakage in OLF patients, with an incidence of 78.8% (6). Complications related to CSF leakage, including pseudocyst, central nervous system infection, and wound dehiscence were common and might lead to catastrophic events if they were not efficiently managed (6,7). Miyakoshi et al. (8) reported that the incidence of dural tears and CSF leaks was higher in patients with DO and dural adhesions, which represented a deleterious factor for preoperative and short-term postoperative neurological status. Muthukumar et al. (9) analyzed the prognostic implications of DO in OLF patients and concluded that DO was not only a detrimental factor for surgical complications but also for prognosis. Additionally, a systematic review by Osman et al. (5) found that dural tears were associated with longer hospital stay, and considered as an independent driver of cost. Given the clinical and economic ramifications of dural tears, surgeons aim to accurately identify the presence of DO before the operation and prepare for the treatment of dural tear.

The incidence of DO in OLF patients varied from 11–62%. Aizawa et al. (10) reported that nine out of 72 OLF patients had dural tears and eight of those had a DO. Miyakoshi et al. (8) reported a higher incidence of dural adhesion in 34 patients (62%), and found a proportionate increase with the severity of OLF. At our institution, the incidence of DO ranged from 25.2–39% (11,12), which was comparable to previous reports. These results indicated that DO was a common complication of OLF, and the diagnosis of DO should be considered in the management of OLF patients.

Currently, several methods are used to diagnose DO, and each method has its own disadvantages in terms of diagnostic accuracy and feasibility. In 2009, Muthukumar et al. (9) first described the radiological characteristics of DO and proposed two types of radiological signs: tram track sign (TTs) and comma sign (Cs). However, due to the relatively small number of patients in the study, the lack of understanding of these specific imaging signs made it impossible to obtain an accurate diagnosis of DO. Sun et al. (7) reported that the diagnostic specificity of TTs was only 59%, and the combination of other methods might be helpful for effective diagnosis. To improve the diagnostic accuracy of DO, Li et al. (11) described a "bridge sign" (Bs), defined and excluded four types of false TTs, and then used TTs, Cs and Bs for combined diagnosis, with sensitivity and specificity of 94.23% and 94.21%, respectively. Zhou et al. (12) analyzed the risk factors of DO and proposed an image grading system to predict DO, with a sensitivity of 76.0% and specificity of 91.0%. The results of these studies indicated that combined use of three or more imaging signs or parameters could significantly improve the diagnostic accuracy. However, no valuable imaging signs have been proposed and validated. Recently, Prasad et al. (13) reported that MRI-T2 ring sign was highly correlated with intraoperative DO (90% sensitivity, 100% specificity) in a small number of patients, and was beneficial to surgeons for preoperative diagnosis of DO. However, this is a newly proposed MRI-based imaging sign, and its accuracy and feasibility need further study.

As one of the largest research centers for thoracic OLF in China, we treat more than 100 OLF patients every year. In clinical practice, we have found a new typical imaging sign, which is considered to be a specific sign of DO, namely "Banner cloud sign" (BCs). This sign is named BCs because the morphology of DO on the sagittal CT reconstruction is similar to the natural landscape of the Banner clouds on mountain peaks (14) (*Figure 1*). The effectiveness and accuracy of BCs in the diagnosis of DO have been preliminarily verified in our daily clinical practice. However, statistical confirmation is lacking. Therefore, we designed this prospective, blinded and diagnostic accuracy study to evaluate the diagnostic accuracy of BCs compare to TTs and Cs, in order to explore the critical role of BCs in the diagnosis of DO.

Materials and methods

Study design

This is a prospective, outcome assessor blinded, diagnostic accuracy study. This study will evaluate and compare the accuracy of BCs, TTs and Cs in the diagnosis of DO in a consecutive series of OLF patients at our center. Patients' medical records and imaging data will be retrieved from the
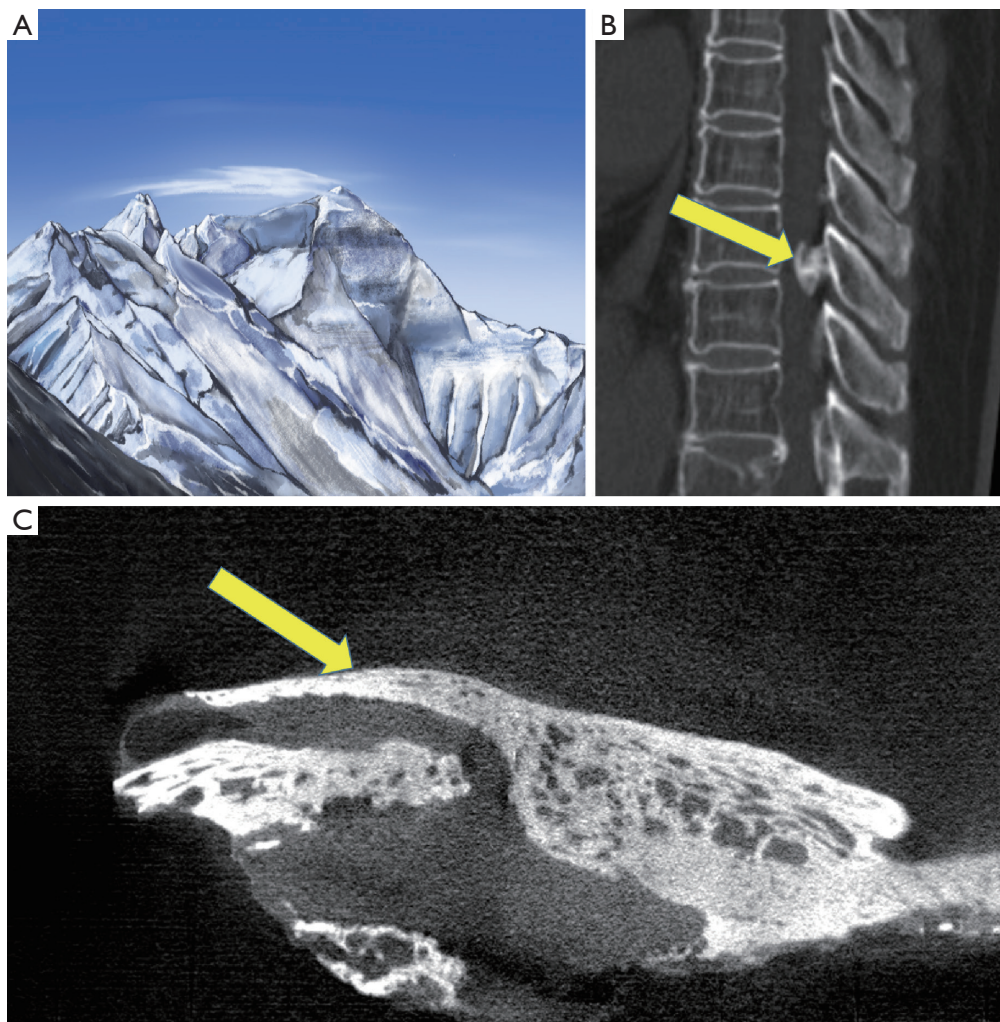
Figure 1 Definition of Banner cloud sign. (A) Natural landscape of the Banner Cloud on mountain peak. (B) A typical imaging sign of dural ossification on sagittal CT reconstruction. (C) Micro-CT scanning illustrates the morphology of dural ossification. Yellow arrow indicates the ossified dura mater.

hospital database server. An observation group comprising of six spine surgeons with different seniority levels and two epidemiological researchers will read the patients' images to identify typical imaging signs and determine the presence of DO. After imaging evaluation, surgical records will be reviewed to confirm the presence of DO, and the results will serve as the reference standard for estimating accuracy. The study workflow is presented in *Figure 2*.

Eligibility criteria

Inclusion criteria
After reviewing the medical records and imaging data,

patients who meet the following criteria will be eligible:
- ❖ Patients with thoracic OLF undergoing posterior decompression surgery;
- ❖ The operation was conducted between January 2018 and June 2019;
- ❖ Complete medical records and operation notes, which can be used to determine the presence of DO;
- ❖ Willing to sign informed consent form.

Exclusion criteria
The exclusion criteria are as follows:
- ❖ OLF patients with thoracic trauma, infection, tumor or deformity;

Page 4 of 8

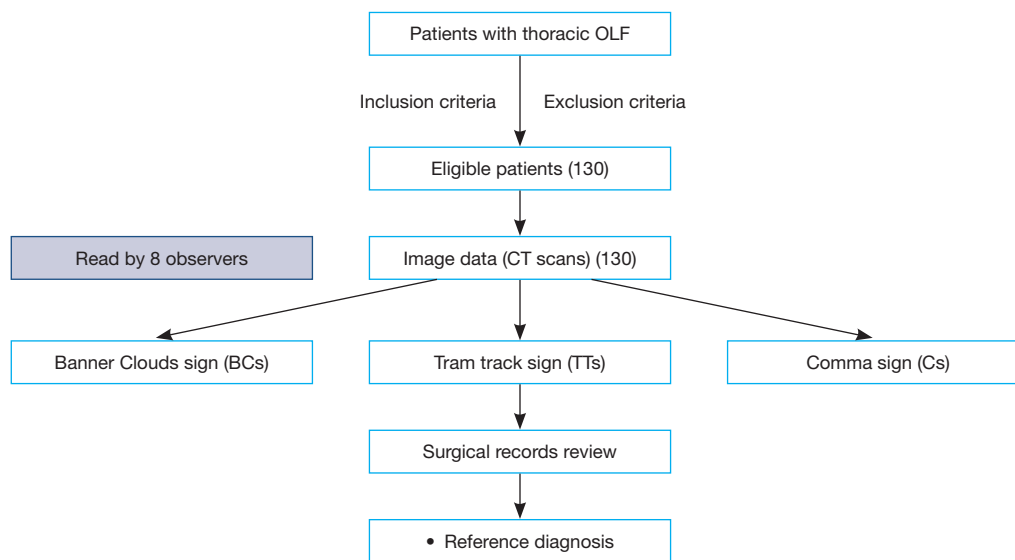Chen et al. CT-based "Banner cloud sign" for DO

Figure 2 Schematic diagram of the study workflow.

❖ OLF combined with diffuse idiopathic skeletal hyperostosis (DISH), Scheuermann's disease, ankylosing spondylitis (AS), skeletal fluorosis or severe rheumatism;
❖ Unwilling to sign informed consent.

Study population

Patients diagnosed with thoracic OLF who received surgical treatment between January 2018 and June 2019 at the Peking University Third Hospital will be recruited. Two surgeons will identify eligible patients based on the medical records and imaging data. Eligible patients will be informed about the study and invited to participate.

Interventions

Two doctors, who are not observers, will be responsible for imaging data collection. Subsequently, statisticians will randomly number all cases in Excel sheets. Before reading the images, all observers will be required to undergo unified training. The principal investigator (PI) will elaborate the typical imaging features of BCs, TTs and Cs in the form of a PowerPoint presentation (PPT), so that all observers are familiar with these three typical signs. Three training sessions at an interval of one week will be conducted. An observation group including six spine surgeons with different seniority levels (two senior titles, two intermediate

titles, two residents) and two epidemiologists with no experience in spine surgery will read the images according to the imaging features of BCs, TTs and Cs to determine the presence of DO.

Outcome measurements

Primary outcome measurements
Since the main purpose of this study is to determine the accuracy of BCs in the diagnosis of DO, the primary outcome measurements are to calculate the sensitivity, specificity, positive and negative predictive values of BCs in the diagnosis of DO, as well as the area under the receiver operating characteristic (ROC) curve.

Secondary outcome measurements
To further evaluate the diagnostic value of BCs, we will calculate and compare it with TTs and Cs based on the sensitivity, specificity, positive predictive value, negative predictive value, Youden's index, likelihood ratio and the area under the ROC curve. As the inter-observer reliability is a critical index to test inter-observer agreement, we will compare the intra-class correlation coefficient (ICC) and Kappa (κ) values of each imaging sign. Since an ideal diagnostic method should be universal and can be mastered by clinicians with different experience levels, the time taken and the level of confidence of using BCs, TTs and Cs for DO diagnosis will also be compared.

Blinding

For the purpose of quality control, strict blinding strategy is necessary. In the reading process, all observers will be blinded to the image reports and any other medical records that can be used to identify the existence of DO.

Data collection and management

We will use the electronic data capture system (EDC) for data collection and management. Each observer will be assigned a separate identification (ID) number as the unique identity to log into the EDC. After logging into the system, the observers will randomly receive the patients' image number. With this number, the observers can access the hospital picture and archiving system (PACS) to retrieve the patients' image data, read them and directly upload the results onto the data management system. A brief description of the data collection form is shown in *Figure 3*. Two statisticians will be responsible for supervising, collating and managing the collected data.

Reference standard

The surgical records will be used as the reference standard to confirm the presence of DO.

Data analysis

Data analysis will primarily focus on the diagnostic accuracy of BCs in DO diagnosis. Sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios will be calculated with corresponding 95% confidence intervals (CIs), by comparing the results of BCs read by eight observers, with the reference diagnosis in the surgical records.

To compare the diagnostic accuracy of different imaging signs, we will calculate the sensitivity, specificity, positive and negative predictive values, Youden index, likelihood ratios and area under the ROC curve of each sign in detecting DO.

To evaluate and compare inter-observer agreement, we will calculate ICCs and κ values of each imaging sign in the diagnosis of DO.

To assess the ease of mastering different imaging signs, we will calculate the time it takes each observer to read the images. The level of confidence of each observer in identifying each imaging sign will be recorded using the five-point Likert scale (none, mild, moderate, very, extreme). Differences between each observer as well as each imaging sign will be tested for statistical significance.

Sample size calculation

We will calculate the sample size based on the area under the ROC curve. Our preliminary experimental results indicated that the AUC of BCs was 0.85. We assume that the AUC of TTs and Cs as the control group is 0.65. Our previous studies reported that the incidence of DO was 25.2–39% (6,7). Therefore, a conservative estimated incidence of 25% will be guaranteed, and the ratio of sample sizes in the positive/negative groups should be 1:3. With the type I error α=0.05 and type II error β=0.1, all the above relevant data will be entered into PASS 14.0 software to calculate the sample size, and a total of 96 patients (24 patients with DO and 72 without DO) will be required. We assume that in a consecutive series of patients, 20% will not meet the inclusion criteria, so a total of 120 patients will be recruited.

Statistical methods

The research data will be managed using Epidata 3.1 software. After the data are exported, SPSS 25.0 software will be used for statistical analysis. The quantitative data in accordance with the normal distribution will be statistically described by the mean ± standard deviation, while those that do not conform to the normal distribution will be described by the median (25%, 75%), and the counting data will be described by the number of cases (%). The comparison of quantitative data will be conducted by independent sample *t*-test, and the comparison of counting data will be conducted by chi-square test or Fisher's exact test. Two-sided P value <0.05 will be defined as statistically significant, with two-sided 90% CIs.

Quality control

For quality control, all observers will be required to undergo unified training prior to the clinical trial. The PI will elaborate the typical imaging features of BCs, TTs and Cs in the form of PPT, and ensure that all observers are familiar with these three typical signs. Three training sessions will be conducted at an interval of one week. To minimize the interference between different imaging signs in the process of image reading, which may produce

Page 6 of 8

Chen et al. CT-based "Banner cloud sign" for DO

Observer ID: XXX											
Random No.		1	2	3	4	5	...	...	..	129	130
Image ID		2418955	xxxx	xxxx	xxxx	xxxx	...	...	..	xxxx	xxxx
Tram track sign	Yes										
	No										
How certain are you about the diagnosis?	Time										
	1 Uncertain 2 Slightly certain 3 Moderately certain 4 Very certain 5 Extremely certain										
re-Random No.		1	2	3	4	5	...	...	..	129	130
Image ID		xxxx	xxxx	xxxx	xxxx	xxxx	...	2418955	..	xxxx	xxxx
Comma sign	Yes										
	No										
How certain are you about the diagnosis?	Time										
	1 Uncertain 2 Slightly certain 3 Moderately certain 4 Very certain 5 Extremely certain										
re-Random No.		1	2	3	4	5	...	...	..	129	130
Image ID		xxxx	xxxx	xxxx	xxxx	xxxx	...	...	..	2418955	xxxx
Banner clouds sign	Yes										
	No										
How certain are you about the diagnosis?	Time										
	1 Uncertain 2 Slightly certain 3 Moderately certain 4 Very certain 5 Extremely certain										

Figure 3 A brief description of the data collection form.

psychological cue effect and affect the judgment of the result, the observers will only judge one type of image sign in each reading process. To guarantee the quality of the entire trial, rigorous monitoring will be performed by three trained quality supervisors. During the trial, supervisors will check the data entry of observers once a week to ensure that the data are true and valid.

Ethics

This study will be conducted in compliance with the principles of the Declaration of Helsinki for Clinical Research. The trial protocol was reviewed and approved by the Research Ethical Committee of the Peking University Third Hospital. Informed consent will be obtained from all participants included in the study. The protocol has been registered on the Clinical Trials website (Trial number is ChiCTR2000030380).

Discussion

OLF is the major cause of thoracic spinal stenosis and has been frequently reported in East Asian countries. With the accumulation of clinical experience and the improvement of diagnostic tools, OLF is being increasingly recognized in Caucasians. In view of the progressive natural course, OLF responds poorly to conservative treatment and surgical intervention is the only effective method when the spinal

cord is severely compressed and neurological impairment occurs. However, in spite of the advances in surgical techniques and instruments, surgery-related complication rate remains high, especially in the case of DO fused with OLF. The involvement of dura membrane in ossification increases the difficulty of surgery, and significantly increases the risks of spinal cord damage and complications. Therefore, there is an urgent need for surgeons to accurately diagnose DO preoperatively and make necessary preparations for the management of intraoperative dural laceration.

There are limited methods available to identify DO, and specific imaging signs based on CT or MRI evidence are most commonly used. However, due to the lack of sample size and limited understanding of the characteristics of DO, the diagnostic accuracy of these existing imaging signs needs to be further improved. In addition, because of different experience levels of doctors and their understanding of typical image features, there is a large intra-group difference in the process of reading images, which may affect the accuracy and consistency of diagnosis results. Theoretically, an ideal diagnostic imaging sign is not only highly specific for DO but also easy to master by surgeons of different seniority levels, so as to improve the diagnostic accuracy.

To the best of our knowledge, this is the first proposed study with a large sample size to evaluate the diagnostic accuracy of BCs and compare it with the diagnostic value of TTs and Cs. Our primary objective is to assess the diagnostic accuracy of BCs for DO by comparing the reading results of multiple observers with the reference standard, and to preliminarily understand the value of BCs in DO diagnosis. To further confirm the diagnostic value of BCs, we will compare BCs with TTs and Cs based on a series of diagnostic indexes such as sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios and area under the ROC curve. If these results are superior to TTs and Cs, then we can conclude that BCs has a higher diagnostic value and can replace previously reported diagnostic methods.

The optimal diagnostic method should not only have high diagnostic accuracy but also be universal, which is easy to understand by observers with different experiences levels. Therefore, our secondary objective is to compare the inter-observer reliability of different methods using ICC and κ values. Inter-observer reliability will be determined by comparing the initial responses of all eight observers. ICC and κ values will be interpreted as follows: 0.00–0.20 indicate slight agreement; 0.21–0.40 indicate fair agreement; 0.41–0.60 indicate moderate agreement; 0.61–0.80 indicate

substantial agreement; and 0.81–1.00 indicate almost perfect agreement (15,16). Additionally, to further evaluate the universality of each imaging sign among clinicians with different seniority levels, we will compare the time required to read the images between observers and between groups, as well as the level of confidence in diagnostic accuracy. If the results of these two indexes show no statistical difference among different observers, then we can conclude that this imaging sign applies to all clinicians regardless of their experience. In contrast, if there are statistical differences in either the time or the level of confidence between different observers or imaging signs, it indicates that these imaging signs have different clinical application value, and the possible differences should be interpreted and analyzed according to the specific situation.

We hope that this study can provide adequate assessment of the diagnostic value of each imaging sign in the diagnosis of DO, with the maximum number of patients. Quality control is crucial to the overall study. To perform a reliable study, we will conduct unified training to ensure that all observers involved in this study fully understand the characteristics of each imaging sign before the start of the study. All observers will be blinded to the identity of the patients and an observer will be allowed to read only one imaging sign at a time, which will further ensure the quality of the study. We hope that this prospective diagnostic accuracy study will increase clinicians' knowledge of the value of each imaging sign in diagnosing DO and provide reliable evidence-based data for their application in clinical practice. The novel imaging sign, BCs may significantly improve the diagnostic accuracy of DO preoperatively, and may aid clinicians to make adequate preparations for the intraoperative management of DO.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-5439). The authors have no conflicts of interest to declare.

Page 8 of 8

Chen et al. CT-based "Banner cloud sign" for DO

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. This study will be conducted in compliance with the principles of the Declaration of Helsinki for Clinical Research (as revised in 2013). The study design, procedure and informed consent procedure were approved by the Peking University Third Hospital (IRB00006761-M2019494). Informed consent will be obtained from all participants included in the study. The protocol has been registered on the Clinical Trials website (Trial number is ChiCTR2000030380).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Guo JJ, Luk KDK, Karppinen J, et al. Prevalence, distribution, and morphology of ossification of the ligamentum flavum: a population study of one thousand seven hundred thirty-six magnetic resonance imaging scans. Spine 2010;35:51-6.
2. Lang N, Yuan H, Wang H, et al. Epidemiological survey of ossification of the ligamentum flavum in thoracic spine: CT imaging observation of 993 cases. Eur Spine J 2013;22:857-62.
3. Chen XQ, Yang HL, Wang GL, et al. Surgery for thoracic myelopathy caused by ossification of the ligamentum flavum. J Clin Neurosci 2009;16:1316-20.
4. Park DA, Kim SW, Lee SM, et al. Symptomatic myelopathy caused by ossification of the yellow ligament. Korean J Spine 2012;9:348-51.
5. Osman NS, Cheung ZB, Hussain AK, et al. Outcomes and Complications Following Laminectomy Alone for Thoracic Myelopathy due to Ossified Ligamentum Flavum: A Systematic Review and Meta-Analysis. Spine 2018;43:E842-8.
6. Sun X, Sun C, Liu X, et al. 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 2012;37:E702-7.
7. Sun XZ, Chen ZQ, Qi Q, et al. Diagnosis and treatment of ossification of the ligamentum flavum associated with dural ossification Clinical article. J Neurosurg Spine 2011;15:386-92.
8. Miyakoshi N, Shimada Y, Suzuki T, et al. 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.
9. Muthukumar N. Dural ossification in ossification of the ligamentum flavum: a preliminary report. Spine 2009;34:2654-61.
10. Aizawa T, Satao T, Sasaki H, et al. Thoracic myelopathy caused by ossification of the ligamentum flavum: clinical features and surgical results in the Japanese population. Journal of neurosurgery. J Neurosurg Spine 2006;5:514-9.
11. Li B, Qiu G, Guo S, et al. Dural ossification associated with ossification of ligamentum flavum in the thoracic spine: a retrospective analysis. BMJ open 2016;6:e013887.
12. Zhou SY, Yuan B, Chen XS, et al. Imaging grading system for the diagnosis of dural ossification based on 102 segments of TOLF CT bone-window data. Sci Rep 2017;7:2983.
13. 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.
14. Schween J, Kuettner J, Reinert D, et al. Definition of "banner clouds" based on time lapse movies. Atmos Chem Phys 2007;7:2047-55.
15. Yu L, Li B, Yu Y, et al. The Relationship between Dural Ossification and Spinal Stenosis in Thoracic Ossification of the Ligamentum Flavum. J Bone Joint Surg Am 2019;101:606-12.
16. Xu C, Yin M, Sun Z, et al. An Independent Inter-observer Reliability and Intra-observer Reproducibility Evaluation of SINS Scoring and Kostuik Classification Systems for Spinal Tumor. World Neurosurg 2020;137:e564-9.