

Independent Reliability Analysis of a New Classification for Pyogenic Spondylodiscitis

Gaston Camino Willhuber, MD^{1,2} , Alfredo Guiroy, MD^{2,3} ,
Juan Zamorano, MD^{2,4} , Nelson Astur, MD^{2,5,6}, and Marcelo Valacco, MD^{2,7}

Global Spine Journal
2021, Vol. 11(5) 669-673
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2192568220919091
journals.sagepub.com/home/gsj



Abstract

Study Design: Diagnostic study, level of evidence III.

Objective: Pyogenic spondylodiscitis can cause deformity, neurological compromise, disability, and death. Recently, a new classification of spondylodiscitis based on magnetic resonance imaging was published. The objective of this study is to perform an independent reliability analysis of this new classification.

Methods: We selected 35 cases from our database of different spine centers in Latin America and from the literature; 8 observers evaluated the classification and graded the scenarios according to the methodological grading of the classification developed by Pola et al. Cases were sent to the observers in a random sequence after 3 weeks to assess intraobserver reliability. The interobserver and intraobserver reliabilities were performed with Fleiss and Cohen statistics, respectively.

Results: The overall Fleiss κ value for interobserver agreement was substantial, with 0.67 (95% CI = 0.43-0.91) in the first reading and 0.67 (95% CI = 0.45-0.89) in second reading for the main types of classification. The Cohen κ value for intraobserver agreement was also substantial, with 0.68 (95% CI = 0.45-0.92). The interobserver agreement analysis for the subtypes of this classification was overall substantial, with 0.60 (95% CI = 0.37-0.83) in the first reading and 0.61 (95% CI = 0.41-0.81) in the second reading. The overall intraobserver agreement for subtypes of the classification was also substantial, with 0.63 (95% CI = 0.34-0.93).

Conclusion: The new classification developed by Pola et al showed substantial interobserver and intraobserver agreements. More studies are required to validate the usefulness of this classification especially in clinical practice.

Keywords

pyogenic spondylodiscitis, spondylodiscitis classification, disc space infection, vertebral osteomyelitis

Introduction

Pyogenic spondylodiscitis (PS) is an infectious disease that involves the vertebral endplates and can extend into the disc space. PS has an estimated prevalence of 6.5 per 100 000 in western societies and is associated with increased morbidity, hospital length of stay, and mortality.¹ This condition has been shown to be more prevalent among elderly patients with chronic debilitating conditions, those with immunodeficiency, and intravenous drug users.² The most common location is the lumbar spine, followed by thoracic and cervical regions,³ and the most frequent agents are staphylococcal sp. and streptococcal species.⁴

The goals of treatment are to relieve pain, avoid neurological deterioration, eradicate infection, provide spinal stability, and prevent deformity.⁵⁻⁷ Orthopedic guidelines with proper

algorithms for the management of PS are not universally accepted and usually rely on clinical studies with variable inclusion criteria.⁸⁻¹² Ideally, a classification system should

¹ Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

² AOSpine Latin America, Curitiba, Brasil

³ Hospital Español, Mendoza, Argentina

⁴ Clínica Alemana de Santiago-Universidad del Desarrollo, Vitacura, Santiago, Chile

⁵ Santa Casa de Misericórdia de São Paulo, São Paulo, Brazil

⁶ Hospital Israelita Albert Einstein, Morumbi, São Paulo, Brazil

⁷ Hospital Churruca Visca, Buenos Aires, Argentina

Corresponding Author:

Gaston Camino Willhuber, Hospital Italiano de Buenos Aires, Potosí 4215, Institute of Orthopedics Buenos Aires, Buenos Aires, C1199ABD, Argentina. Email: gaston.camino@hospitalitaliano.org.ar



Table 1. Classification of Spondylodiscitis According to Pola et al.¹⁷

Type	Description
Type A	All cases without biomechanical instability, epidural abscesses, or neurological involvement
A1	Simple discitis without the involvement of vertebral bodies
A2	Spondylodiscitis involving the intervertebral disc and adjacent vertebral bodies
A3	Spondylodiscitis with limited involvement of paravertebral soft tissues
A4	Spondylodiscitis with unilateral (A.4.1) or bilateral (A.4.2) intramuscular abscesses
Type B	Includes cases with radiological instability of significant bone destruction without epidural abscesses or neurological involvement
B1	Destructive spondylodiscitis without segmental instability
B2	Destructive spondylodiscitis extended to paravertebral soft tissues without segmental instability
B3	Destructive spondylodiscitis with biomechanical instability and segmental kyphosis
Type C	All cases with neurological compromise or epidural abscesses
C1	Epidural abscess without neurological symptoms neither segmental instability
C2	Epidural abscess and segmental instability without neurological impairment
C3	Epidural abscess and acute neurological impairment without segmental instability
C4	Epidural abscess and acute neurological impairment with segmental instability

be easy to apply, inclusive, reproducible,^{13,14} and if possible, helpful in guiding a treatment option with recommendations.

Previous attempts to create a proper classification system with treatment guidelines has been proposed^{15,16}; however, to date, no classification system has been universally accepted. In 2017, Pola et al¹⁷ developed a new classification of PS based on contrast-enhanced magnetic resonance imaging (MRI) findings to define a treatment algorithm. The objective of this study is to perform a reliability study of the new classification developed by Pola et al¹⁷ through interobserver and intraobserver analyses to assess the concordance of this classification among readers.

Material and Methods

After approval from our institutional review board (Protocol Number IRB 0 003 937), we conducted a multicenter study to assess a validation through an independent analysis of the classification. We identified 8 young spine surgeons (observers) to evaluate 35 PS contrast-enhanced MRI scenarios each one related to different spondylodiscitis cases according to the classification described by Pola et al¹⁷ (see Table 1). In their study, Pola et al¹⁷ described the clinical-radiological classification of spondylodiscitis in 250 patients with treatment recommendations and at least a 2-year follow-up. The classification is based on MRI according to major criteria, such as the presence of instability, epidural abscess, and neurological compromise, and minor criteria, such as paravertebral soft tissue of intramuscular abscess. The authors also proposed a

Table 2. Treatment Algorithms According to the Classification.

Classes	Treatments of Choice
Type A	
A1	Rigid orthosis immobilization
A2-A4	Rigid orthosis immobilization or percutaneous stabilization
Type B	
B1-B2	Rigid orthosis immobilization or percutaneous stabilization
B3	Percutaneous or open stabilization
Type C	
C1	Rigid orthosis immobilization or percutaneous stabilization with closer clinical-radiological monitoring
C2	Open debridement and stabilization
C3	Open debridement and decompression
C4	Open debridement, decompression, and stabilization

Table 3. Distribution of Pyogenic Spondylodiscitis Cases According to the Main Types.

Type	Number	Percentage
Type A	12	34
Type B	11	32
Type C	12	34
Overall	35	100

treatment algorithm based on the main types and subtypes (Table 2).

The clinical scenarios were gathered from a database of 6 centers in Latin America by the authors of the study. The invited observers were spine fellows from diverse centers in Latin America who did not belong to the designer team and were not familiarized with the included cases. The designer team (authors) received 97 cases of spondylodiscitis; each case was received by an email containing a case presentation with MRI and clinical information (presence of neurological compromise based on physical examination). All cases were classified individually by the authors according to the classification developed by Pola et al.¹⁷ This classification consists of 3 main types (A, B, and C) based on the presence of primary criteria on MRI: bone destruction of segmental instability, epidural abscess, and neurological impairment. Secondary criteria help define subtypes of the classification and are as follows: soft-tissue involvement and paravertebral muscular abscesses. Cases reaching a concordance of 100% among the authors were subselected, and 35 cases were then finally selected to be included in the analysis. The classification was sent by email to the observers; each observer was previously trained to apply the classification; and questions were explained before the final assessment. All 35 clinical cases (12 type A, 11 type B, and 12 type C cases; Table 3) were sent at once to the observers by email, and they had access to the classification scheme while grading (Table 1). All cases were sent back to the designer team with each respective classification type. After 3 weeks, all participating appraisers received the same 35 cases again, but in a different



Figure 1. Cases of spondylodiscitis: (A, B) Type A spondylodiscitis. (C, D) Type B spondylodiscitis. (E, F) Type C Spondylodiscitis.

random order, to classify and send results back to the authors for intraobserver reliability analysis.

All data was collected and analyzed for reliability. Interobserver and intraobserver agreements were assessed in 2 different ways: for the main type of classification (types A, B and C) and for each subtype regarding complication (A1, A2, B2, etc).

Treatment Management

Our 35 cases were treated similarly according to the algorithm by Pola et al¹⁷ (Table 2). Of the 12 type A cases, 10 received conservative treatment with orthosis and 2 cases required stabilization for persistent back pain. Four type B cases were treated conservatively, and 7 required stabilization; 8 type C cases required open debridement and decompression, and 4 cases required stabilization after decompression.

Statistical Analysis

Evaluation of Interobserver Agreement. In a first step, we evaluated the interobserver agreement through the calculation of the unweighted κ coefficient for the main type classification. The

classification for main categories (types) for spondylodiscitis (Figure 1; types A, B, and C) was prepared as an ordinal variable represented by numerical values from 1 to 3 (type A: 1; type B: 2; and type C: 3); the authors assigned 1 point to perfect agreement between each pair of readers when both readers assigned the same type, and perfect disagreement was represented as a value of 0 when different types were assigned by 2 readers.

In a second step, we evaluated the interobserver agreement through the calculation of the weighted κ coefficient for subtypes of classification for each pair of judges (readers). Regarding the classification categories (subtypes) for spondylodiscitis, an ordinal variable represented by numerical values from 1 to 11 was prepared (subtype A1: 1; subtype A2: 2; subtype A3: 3; subtype A4: 4; subtype B1: 5; subtype B2: 6; subtype B3: 7; subtype C1: 8; subtype C2: 9; subtype C3: 10; subtype C4: 11).

The authors assigned 1 point to the perfect agreement between each pair of readers, defined as the situation in which both readers assigned exactly the same subtype of spondylodiscitis classification to the clinical scenario in question. When both readers had a disagreement of more than 1 category, it was represented by a value of 0. When the disagreement was of only 1 category of difference (eg, the same case was evaluated as

Table 4. Interobserver Agreement of Main Types of Spondylodiscitis.

Type	First Reading Fleiss K (95% CI)	Second Reading Fleiss K (95% CI)
Type A	0.67 (0.42-0.89)	0.77 (0.41-0.89)
Type B	0.53 (0.39-0.78)	0.49 (0.38-0.81)
Type C	0.77 (0.51-0.91)	0.76 (0.53-0.87)
Overall	0.67 (0.43-0.91)	0.67 (0.45-0.89)

corresponding to grade 2 of the classification by a judge and to grade 1 by the other reader), the authors agreed on an intermediate penalty represented by 0.5 or 0.75 points according to the case. A minimal penalty was applied to the close disagreements of low clinical relevance, representing their degree of agreement with a value of 0.75; those of greater clinical relevance were given an intermediate penalty, representing their agreement with a value of 0.5.

Evaluation of the Degree of Intraobserver Agreement. To assess the degree of intraobserver agreement (test-retest), a weighted κ coefficient was calculated according to the same weighting matrix as for the degree of agreement among the different observers. We determined sample size to provide adequate variability to assess discrimination among the main types of spondylodiscitis and acceptably precise reliability estimates. Based on a simulation process, when the sample consisted of 35 participants, each being assessed by 8 raters on a 3-category classification system, there would be a greater than 95% chance to reject the null hypothesis that the Fleiss j is less than 0.7; if true, then the Fleiss j is 0.9. Chance-adjusted Fleiss and Cohen statistics with 95% CIs were used to determine interobserver and intraobserver reliabilities, respectively.^{18,19}

The level of agreement (κ) was determined as proposed by Landis and Koch²⁰ with κ values of 0.00 to 0.20 considered slight agreement, 0.21 to 0.40 considered fair agreement, 0.41 to 0.60 considered moderate agreement, 0.61 to 0.80 considered substantial agreement, and 0.81 to 1.00 considered almost perfect agreement.

Results

Reliability for the Main Types of Spondylodiscitis

The overall interobserver agreement was 0.67 (95% CI = 0.43-0.91) for the first reading and 0.67 (95% CI = 0.45-0.89) for second readings. Agreement analysis based on each type is described in Table 4. Intraobserver agreement was 0.68 (95% CI = 0.45-0.92).

Reliability Analysis of the Subtypes of Spondylodiscitis

Assessment of reliability among observers for different subtypes of spondylodiscitis showed an interobserver agreement of 0.60 (95% CI = 0.37-0.83) in the first reading and 0.61 (95% CI = 0.41-0.81) in the second readings. A substantial intraobserver agreement of 0.63 (95% CI = 0.34-0.93) was found.

Discussion

Our study showed substantial interobserver ($\kappa = 0.67$) and intraobserver ($\kappa = 0.68$) agreements of the classification of Pola et al,¹⁷ with moderate ($\kappa = 0.53$ and 0.49) agreement when classifying type B. The main reason that can explain why agreement was lower in type B could be that this classification considers spinal instability (defined as more than 25% in segmental kyphosis at the level compromised) as the primary criterion to differentiate between A and B types. This is a limitation of the classification and may lead to a potential misconception when classifying, especially because spinal instability is not always observed on MRI because this imaging exam is performed in the supine position. Instability criteria are better assessed through a standing radiograph instead of MRI; however, there are cases where instability secondary to bone destruction is evident even on MRI. Furthermore, many patients are unable to maintain a standing position for an X-ray, and we agree with the authors that MRI is the best modality to classify spondylodiscitis.

Another limitation of this classification is that it relies on an extensive number of subtypes, especially considering that different subtypes such as A2, A3, and A4 require the same treatment according to the author's recommendation, with a similar concept in B1 and B2 subtypes. Probably a more simplified classification could be easier to understand and still useful regarding the treatment recommendations.¹⁷ A classification system should be reproducible and useful for widespread acceptance in clinical practice and research; Pola et al¹⁷ conducted no agreement analysis in their study, and to our knowledge, this is the first independent analysis for this classification.

Our study has strengths, such as the number of observers, which makes the results more reliable. In addition, the observers were from different institutions in Latin America and were all actively involved in spine surgery, not limiting the analysis to a single center. Another strength is that this reliability study was conducted by spine surgeons from a region other than Pola et al,¹⁷ which decreases conflicts of interests.

It is important to state the limitations of our study. First, the designer team selected cases from a database and decided which cases were more likely to be analyzed and compared; this could represent a selection bias. On the other hand, 100% concordance among the authors was required to include the case in the analysis, making the selection process more reliable. Another limitation is the level of training of the observers, which included young spine surgeons instead of experienced attending surgeons. This could affect interobserver analyses. Ideally, experienced, trained spine surgeons could yield more reliable results. However, reliability studies usually include fellowship-trained physicians who usually show an active participation and availability for these studies, and many interobserver analyses showed no difference between residents and attending surgeons in terms of agreement.^{13, 21, 22} Another limitation of this study is the lack of interobserver agreement between each subtype of the classification, probably a result of the number of cases required to make this specific measurement; instead, we evaluated the overall

reliability of subtypes among readers, which was substantial (0.61). Knowing the reliability to classify each subtype would be ideal to understand which subtypes of the classification are most difficult to identify; on the other hand, the number of cases required to carry out this estimation is high (more than 100 cases to assess by each observer), and this also can affect the judgment of the observers because of the extensive analysis. Despite these limitations, this is the first independent interobserver and intraobserver agreement analysis to assess reliability of the new classification developed by Pola et al.¹⁷

Conclusion

The new classification of spondylodiscitis proposed by Pola et al,¹⁷ even though it is extensive regarding the subtypes, has shown substantial interobserver and intraobserver agreements and could be an important tool when classifying PS. More studies are required to evaluate the usefulness of this classification, especially in clinical practice.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Gaston Camino Willhuber  <https://orcid.org/0000-0002-5684-7679>
 Alfredo Guiroy  <https://orcid.org/0000-0001-9162-6508>
 Juan Zamorano  <https://orcid.org/0000-0002-7284-0784>

References

- Loibl M, Stoyanov L, Doenitz C, et al. Outcome-related cofactors in 105 cases of vertebral osteomyelitis in a tertiary care hospital. *Infection*. 2014;42:503-510.
- Govender S. Spinal infections. *J Bone Joint Surg Br*. 2005;87:1454-1458.
- Wisneski RJ. Infectious disease of the spine: diagnostic and treatment considerations. *Orthop Clin North Am*. 1991;22:491-501.
- Mann S, Schutze M, Sola S, Piek J. Nonspecific pyogenic spondylodiscitis: clinical manifestations, surgical treatment, and outcome in 24 patients. *Neurosurg Focus*. 2004;17:E3.
- Zarghooni K, Röllinghoff M, Sobottke R, Eysel P. Treatment of spondylodiscitis. *Int Orthop*. 2012;36:405-411. doi:10.1007/s00264-011-1425-1
- Rutges JP, Kempen DH, van Dijk M, Oner FC. Outcome of conservative and surgical treatment of pyogenic spondylodiscitis: a systematic literature review. *Eur Spine J*. 2016;25:983-999. doi:10.1007/s00586-015-4318-y
- Lin CP, Ma HL, Wang ST, Liu CL, Yu WK, Chang MC. Surgical results of long posterior fixation with short fusion in the treatment of pyogenic spondylodiscitis of the thoracic and lumbar spine: a retrospective study. *Spine (Phila Pa 1976)*. 2012;37:E1572-1579. doi:10.1097/BRS.0b013e31827399b8
- Aljawadi A, Jahangir N, Jeelani A, et al. Management of pyogenic spinal infection, review of literature. *J Orthop*. 2019;16:508-512. doi:10.1016/j.jor.2019.08.014
- Korovessis P, Vardakastanis K, Fennema P, Syrimbeis V. Mesh cage for treatment of hematogenous spondylitis and spondylodiscitis. How safe and successful is its use in acute and chronic complicated cases? A systematic review of literature over a decade. *Eur J Orthop Surg Traumatol*. 2016;26:753-761. doi:10.1007/s00590-016-1803-x
- Mavrogenis AF, Megaloikonomos PD, Igoumenou VG, et al. Spondylodiscitis revisited. *EFORT Open Rev*. 2017;2:447-461. doi:10.1302/2058-5241.2.160062
- Shiban E, Janssen I, Meyer B, Reiner J, Ringel F. Long-term outcome following surgical treatment for spondylodiscitis in 211 cases. *Global Spine J*. 2016;6(1, suppl):s-0036-1583123-s-0036-1583123. doi:10.1055/s-0036-1583123
- Sudarshan P, Panda A, Paramasivam A, Varma VK, Hegde S. Early diagnosis and operative management in non tuberculous spondylodiscitis: outcome in a case series of 34 patients. *Global Spine J*. 2016;6(1, suppl):s-0036-1583131-s-0036-1583131. doi:10.1055/s-0036-1583131
- Urrutia J, Zamora T, Yurac R, et al. An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. *Spine (Phila Pa 1976)*. 2015;40:E54-E58. doi:10.1097/BRS.0000000000000656
- Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? *A systematic review of their methodologies*. *Acta Orthop Scand*. 2004;75:184-194.
- Akbar M, Lehner B, Doustdar S, et al. Pyogenic spondylodiscitis of the thoracic and lumbar spine: a new classification and guide for surgical decision-making [in German]. *Orthopade*. 2011;40:614-623. doi:10.1007/s00132-011-1742-5
- Homagk L, Homagk N, Klauss JR, Roehl K, Hofmann GO, Marmelstein D. Spondylodiscitis severity code: scoring system for the classification and treatment of non-specific spondylodiscitis. *Eur Spine J*. 2016;25:1012-1020. doi:10.1007/s00586-015-3936-8
- Pola E, Autore G, Formica VM, et al. New classification for the treatment of pyogenic spondylodiscitis: validation study on a population of 250 patients with a follow-up of 2 years. *Eur Spine J*. 2017;26(suppl 4):479-488. doi:10.1007/s00586-017-5043-5
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
- Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378-382.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:158-174.
- Palma J, Villa A, Mery P, et al. A new classification system for pilon fractures based on CT scan: an independent interobserver and intraobserver agreement evaluation. *J Am Acad Orthop Surg*. 2020;28:208-213. doi:10.5435/JAAOS-D-19-00390
- Fox S, Spiess M, Hnenny L, Fournery DR. Spinal Instability Neoplastic Score (SINS): reliability among spine fellows and resident physicians in orthopedic surgery and neurosurgery. *Global Spine J*. 2017;7:744-748. doi:10.1177/2192568217697691