Reliability Statistics: A “Weight-Bearing” View

Aeidin Abedi, MD¹, Nassim Lashkari, PharmD¹, Zorica Buser, PhD¹, and Jeffrey C. Wang, MD¹

Dear Editor,

We read with interest the manuscript by Dettori and Norvell, in which they addressed some of the limitations of kappa statistics and discussed the application of Gwet’s Agreement Coefficient (AC₁) as an alternative.¹ The extent to which the choice of statistical methods or “benchmark” can affect the reliability analysis is indeed surprising. The paradoxical behavior of the kappa statistics seems to occur when the trait prevalence is at extremes, although its confounding effect can be minimal when sensitivity approximates specificity.² Types of measurements utilized in clinical research and practice are broad, particularly in our field of spine. Physical examinations are part of the physicians’ daily practice, imaging classifications are an indispensable part of Reporting and Data Systems, and patient-reported outcome measures are gaining increasing interest in face of the patient-centered care paradigm. Therefore, we would like to commend the authors for their concise discussion and supplement their work by addressing another caveat of reliability studies that we often encounter in the literature.

For the purpose of this communication and to further our understanding about this topic, we reviewed a random sample of 50 studies published in the Global Spine Journal that included one or more categorical variables, such as a classification, grading, staging, or scoring measure. Of these, 24 studies included reliability testing within their statistical analysis (Table 1), among which 17 (71%) used the kappa statistics, 4 (17%) used the intraclass correlation coefficient (ICC), 2 (8%) used both the kappa and ICC, and 1 (4%) used both Cronbach’s α and ICC analysis. Based on our experience, a notable proportion of the available reliability studies suffer from a methodological limitation due to the use of unweighted coefficients.³ The rationale behind the use of weighted method is that “[...] some disagreements in assignments [...] are of greater gravity than others”⁴; hence, in some situations the extent of disagreement and the distance between the categories should be taken into account. For instance, in assessment of spinal instability in patients with metastatic tumors, misclassification of stable patients as “unstable” would pose the risk of unnecessary surgery, while misclassification to “impending instability” has less dire consequences for this vulnerable population. The use of weighted kappa is among the methodological standards for quality assessment of studies on reliability of ordinal scores.⁴,⁵ Fortunately, the weighted extension of Gwet’s AC₁ method is also available and has been described as the AC₂ coefficient.⁶ Overall, we would recommend the use of weighted coefficients to be considered when reliability studies are designed. While we agree that the AC₁ or AC₂ coefficients might be good alternatives, we believe that the use of “weighted” kappa—when applicable—would still be a good progress toward methodologically sound assessments of reliability.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Disclosures outside of submitted work: JCW- Royalties—Biomet, Seaspine, Amedica, DePuy Synthes; Investments/Options—Bone Biologics, Pearldiver, Electrocore, Surgitec; Board of Directors - North American Spine Society, AO Foundation (20,000 honorariums for board position, plus travel for board meetings), Cervical Spine Research Society; Editorial Boards - Spine, The Spine Journal, Clinical Spine Surgery, Global Spine Journal; Fellowship Funding (paid directly to institution): AO Foundation. ZB- consultancy: Cerapedics, The Scripps Research Institute, Xencos Medical (past), AO Spine (past); Research Support: SeaSpine (past, paid to the institution), Next

¹ Department of Orthopaedic Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Corresponding Author:
Zorica Buser, Department of Orthopaedic Surgery, Keck School of Medicine, University of Southern California, 1430 San Pablo St. HC4—#5400A, Los Angeles, CA 90033, USA.

Email: zbuser@usc.edu

Creative Commons Non Commercial No Derives CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerives 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Table 1. Presence or absence and type of reliability analysis in included studies.

| Article | Study design | Objective of study and grading/classification methods | Presence and type of reliability analysis |
|---------|--------------|------------------------------------------------------|------------------------------------------|
| Suzuki A, Daubs MD, Hayashi T, et al. Patterns of cervical disc degeneration: analysis of magnetic resonance imaging of over 1000 symptomatic subjects. *Global Spine Journal*. 2018;8(3):254-259. doi:10.1177/192568217719436 | Cross-sectional | The authors developed a comprehensive grading system and an algorithm for cervical disc degeneration (Table 1). The grade of cervical disc degeneration was categorized into 4 grades: Grade 0 (no degeneration), Grade I (mild degeneration), Grade II (moderate degeneration), Grade III (severe degeneration). | Yes, kappa statistic |
| Garber S, Bisson E, Schmidt M. Comparison of three-dimensional fluoroscopy versus postoperative computed tomography for the assessment of accurate screw placement after instrumented spine surgery. *Global Spine Journal*. 2012;2(2):095-098. doi:10.1055/s-0032-1319775 | Retrospective | The objective of study was to compare images obtained using 3-dimensional fluoroscopy intraoperatively with the images obtained on postoperative CT scan for equivalent visualization of instrumentation placement. | Yes, kappa statistic |
| Dahdaleh N, Khanna R, Menezes A, et al. The application of the revised condyle-C1 interval method to diagnose traumatic atlanto-occipital dissociation in adults. *Global Spine Journal*. 2016;6(6):529-534. doi:10.1055/s-0035-1569058 | Retrospective | Modified or revised condyle-C1 interval (CCI) along with other commonly used methodologies were retrospectively applied to patients with atlanto-occipital dislocation (AOD) and to a cohort of 30 patients without AOD. | Yes, kappa statistic |
| Malham G, Parker R, Ballok Z, Goss B, Diwan A, Uribé J. Bone scans are reliable for the identification of lumbar disk and facet pathology. *Global Spine Journal*. 2015;5(1):023-030. doi:10.1055/s-0034-1394298 | Survey | The objective of the study was to evaluate the reliability of bone single-photon emission computed tomography (SPECT) versus bone SPECT images co-registered with computed tomography (bone SPECT-CT). | Yes, kappa statistic |
| Kulkarni AG, Tapashetti S. Outcomes of discectomy in young adults with large central lumbar disc herniations presenting with predominant leg pain. *Global Spine Journal*. 2019;10(4):412-418. doi:10.1177/192568219856871 | Retrospective cohort | The objective of the study was to evaluate the outcomes of discectomy in young patients with large central lumbar disc herniation (CLDH) presenting with predominant leg pain. | Yes, kappa statistic |
| Willhuber GC, Bendersky M, Cicco FLD, et al. Development of a new therapy-oriented classification of intervertebral vacuum phenomenon with evaluation of intra- and interobserver reliabilities. *Global Spine Journal*. 2020;219:256822091300. doi:10.1177/192568220913006 | Diagnostic study | The objective of this study was to describe a new classification of vacuum phenomenon based on computed tomography scan in order to improve the indications for percutaneous discoplasty. | Yes, kappa statistic |
| Barkoh K, Othorhenan E, Lee L, et al. The DOWN questionaire: a novel screening tool for cervical spondylotic myelopathy. *Global Spine Journal*. 2019;9(6):607-612. doi:10.1177/192568218815863 | Case-control | The authors created the DOWN questionnaire to identify patients with cervical spondylotic myelopathy (CSM) that should be further evaluated both clinically and with advanced imaging. | Yes, kappa statistic |
| Ono AHDA, Chang VYP, Rodenbeck EM, et al. Assessment of the accuracy of the AO Spine-TL classification for thoracolumbar spine fractures using the AO Surgery Reference mobile app. *Global Spine Journal*. 2020;219:256822090169. doi:10.1177/192568220901694 | Cohort | The authors sought to evaluate the accuracy of the AO Surgery Reference mobile app in the diagnosis and treatment of thoracolumbar fractures as proposed by the AO TL classification. | Yes, kappa statistic |
| Willhuber GC, Guiroy A, Zamarano J, Astur N, Valacco M. Independent reliability analysis of a new classification for pyogenic spondylodiscitis. *Global Spine Journal*. 2020;219:256822091909. doi:10.1177/192568220919091 | Diagnostic | The authors sought to perform an independent reliability analysis of a new classification system for spondylodiscitis based on magnetic resonance imaging (Table 1). | Yes, kappa statistic |
| Stoker GE, Buchowski JM, Chen CT, Kim HJ, Park MS, Riew KD. Hypovitaminosis D and cervical disk herniation among adults undergoing spine surgery. *Global Spine Journal*. 2013;3(4):231-236. doi:10.1055/s-0033-1354252 | Retrospective | The authors sought to examine the association of vitamin D levels with intervertebral disk disease. | Yes, kappa statistic |

(continued)
| Article | Study design | Objective of study and grading/classification methods | Presence and type of reliability analysis |
|---------|--------------|-------------------------------------------------------|------------------------------------------|
| Park M, Moon SH, Kim TH, et al. | Retrospective | Asymptomatic spinolaminar line test and ulna classification for assessing cervical stenosis in patients with symptomatic lumbar stenosis. | Yes, kappa statistic |
| Rajasekaran S, Kanna RM, Schroeder GD, et al. | Prospective survey | The authors evaluated how surgeon’s experience affected the reproducibility of fracture classification according to the AOSpine Thoracolumbar Injury Classification System. | Yes, kappa statistic |
| Oshima Y, Kelly M, Song KS, et al. | Retrospective cohort | The objective of the study was to clarify the sensitivity of C3-C2 spinolaminar line test as a screening tool for the stenosis of C1 space available for the cord (SAC). | Yes, kappa statistic |
| Grin A, Krylov V, Lvov I, et al. | Multicenter observational survey | The objective of the study was to measure and compare the intra- and interobserver reliability for the Allen-Fergusson, Harris, Argenson, and AOSpine systems of classification of subaxial cervical spinal injuries, on implementation by neurosurgeons having different levels of experience and working in different clinics. | Yes, kappa statistic |
| Weber C, Rao V, Gulati S, Kivistad K, Nygaard Ø, Lamme G | Validation study | The objective of the study was to evaluate and validate the intra- and interobserver agreement of a morphological grading system for central lumbar spinal stenosis on magnetic resonance imaging between neurosurgeons and radiologists. | Yes, kappa statistic |
| Schroeder G, Kepler C, Koerner J, et al. | Survey | The objective of the study was to determine the reliability with which spine surgeons identify a posterior ligamentous complex (PLC) injury in a patient with a compression-type vertebral body fracture (type A). | Yes, kappa statistic |
| Schell A, Rhee JM, Holtbrook J, Lenehan E, Park KY. | Retrospective radiographic | The objective of the study was to determine interrater variability in grading cervical foraminal stenosis using 3 different CT imaging modalities: 3D CT surface reconstructions (3DSR), 2D sagittal oblique multiplanar reformations (2D-SOMPR), and conventional 2D axial CT imaging. | Yes, kappa statistic |
| Morimoto Y, Shigematsu H, Iwata E, et al. | Retrospective | The authors sought to clarify the differences in cervical alignment findings between sitting cervical lateral radiographs and standing whole-spine lateral radiographs with clavicle positioning in cervical spondylotic myelopathy (CSM) patients. | Yes, intraclass correlation coefficient (ICC) |
| Cheung J, Samartzis D, Cheung P, Cheung K, Luk K. | Prospective radiographic study | The objective of the study was to test the reliability of the Distal Radius and Ulna Classification (DRU) | Yes, intraclass correlation (ICC) analysis |
Table 1. (continued)

| Article                                                                 | Study design      | Objective of study and grading/classification methods                                                                 | Presence and type of reliability analysis |
|------------------------------------------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------|------------------------------------------|
| Abraham E, Manson N, McKeon M. The incidence of adjacent segment breakdown in polysegmental thoracolumbar fusions of three or more levels with minimum 5-year follow-up. Global Spine Journal 2014;4(2):083-088. doi:10.1055/s-0034-1370693 | Retrospective cohort | The objective of the study was to identify the incidence of adjacent segment pathology (ASP) after thoracolumbar fusion of 3 or more levels, the risk factors for the development of ASP. Radiographs exhibiting no adjacent segment pathology were distinguished from radiographs demonstrating findings such as degenerative disk disease, listhesis, instability, stenosis, and/or deformity. | No                                        |
| Horsting P, Pavlov P, Jacobs W, Obradov-Rajic M, de Kleuver M. Good functional outcome and adjacent segment disc quality 10 years after single-level anterior lumbar interbody fusion with posterior fixation. Global Spine Journal 2012;2(1):021-026. doi:10.1055/s-0032-130764 | Prospective cohort | The authors sought to analyze long-term functional outcome and to determine the rate of adjacent segment degeneration (ASD) 10 years after ALIF with additional posterior fixation. | No                                        |
| Obraud M, Bénard MR, Janssen MMA, Anderson PG, Heesterbeek PJ, Spruit M. Kinematic magnetic resonance imaging assessment of the degenerative cervical spine: changes after anterior decompression and cage fusion. Global Spine Journal 2016;6(6):673-678. doi:10.1055/s-0036-1579551 | Prospective cohort | The objective of the study was to measure adjacent segment pathology, cervical curvature, and ROM of individual segments of the cervical spine using kinematic MRI before and 24 months after monosegmental cage fusion. Stage of degeneration, grade of spinal stenosis, and change in cervical curvature was measured before and after monosegmental cage fusion Stage of degenerative disease: “Degenerative cascade,” MRI based classification, consists of 3 stages (I, II, III) Spinal stenosis: Classification of Muhle et al, consists of 4 grades (0-3) Cervical curvature: Guigui et al and by Batzdorf and Batzdorff, consists of type 0-3 | No                                        |
| Grin A, Krylov V, Lykov I, et al. External multicenter study of reliability and reproducibility for lower cervical spine injuries classification systems—part 2: an analysis of the subaxial cervical spine injury classification and cervical spine injury severity score scale. Global Spine Journal 2019;2192568219896546. doi:10.1177/2192568219896546 | Reliability analysis | The authors sought to quantify and compare inter- and intraobserver reliability of the subaxial cervical spine injury classification (CSISS) and the cervical spine injury severity score (CSISS). | Yes, intraclass correlation coefficient (ICC) and kappa statistic |
| Gupta A, Upadhyaya S, Yeung CM, et al. Does size matter? An analysis of the effect of lumbar disc herniation size on the success of nonoperative treatment. Global Spine Journal 2019;2192568219880822 | Retrospective study | Authors examined whether the size of a lumbar disc herniation (LDH) is predictive of the need for surgical intervention within 2 years after obtaining an initial MRI scan. Following measurements were made for a given disc herniation: The anterior-posterior length of both the canal and the herniated disc, the average width of the disc within the canal, the total canal area, and the total disc area. | Yes, Cronbach's alpha and intraclass correlation coefficient (ICC) |
| Assenie SHG, van Hooff ML, Pouw M, de Kleuver M, Spruit M. The natural history of progression in adult spinal deformity: a radiographic analysis. Global Spine Journal. 2020;10(3):272-279.doi:10.1177/2192568219845659 | Historical cohort | The objective of the study was to evaluate the natural history of adult spinal deformity (ASD), and curve progression in the coronal and sagittal planes in a cohort of nonsurgical patients. Radiographic factors measured included direction of scoliosis, curve magnitude, Cobb angle, thoracic kyphosis, lumbar lordosis, pelvic tilt, pelvic incidence, sacral slope, pelvic incidence minus lumbar lordosis and position of the intercrest line. | Yes, intraclass correlation (ICC) |
| Fox S, Spiess M, Hnenny L, Fournier DR. Spinal Instability Neoplastic Score (SINS): reliability among spine fellows and resident physicians in orthopedic surgery and neurosurgery. Global Spine Journal 2017;78(4):744-748. doi:10.1177/2192568217697691 | Pilot study       | The objective of the study was to determine the interobserver variability and the feasibility of CT based parameters to assess posterior ligament complex (PLC) injury compared with MRI when used in a multicenter setting. | Yes, intraclass correlation (ICC) |
| Maheswaran A, Aiyer SN, Farouk O, et al. Assessment of interobserver variability for CT scan–based evaluation of posterior ligament complex injury in thoracolumbar fractures: an international multicenter pilot study. Global Spine Journal. 2020;10(2):118-129.doi:10.1177/2192568219896546 | Pilot study       | The objective of the study was to determine the interobserver variability and the feasibility of CT based parameters to assess posterior ligament complex (PLC) injury compared with MRI when used in a multicenter setting. | Yes, intraclass correlation (ICC) |
| Abraham E, Manson N, McKeon M. The incidence of adjacent segment breakdown in polysegmental thoracolumbar fusions of three or more levels with minimum 5-year follow-up. Global Spine Journal. 2014;4(2):083-088. doi:10.1055/s-0034-1370693 | Retrospective cohort | The objective of the study was to identify the incidence of adjacent segment pathology (ASP) after thoracolumbar fusion of 3 or more levels, the risk factors for the development of ASP. Radiographs exhibiting no adjacent segment pathology were distinguished from radiographs demonstrating findings such as degenerative disk disease, listhesis, instability, stenosis, and/or deformity. | No                                        |
The objective of the study was to investigate whether mechanical bowel preparation ameliorate surgical performance in anterior lumbar interbody fusion. Global Spine Journal. 2019;10(4):393-398. doi:10.1177/2192568219850143

Rasack K, Khoury J, Aoude A, et al. The effect of thoracolumbar pedicle isthmus on pedicle screw accuracy. Global Spine Journal. 2019;10(4):393-398. doi:10.1177/2192568219850143

Jeon CH, Lee HD, Chung NS. Does mechanical bowel preparation ameliorate surgical performance in anterior lumbar interbody fusion? Global Spine Journal. 2019;9(7):692-696. doi:10.1177/2192568218825249

Lee HQ, Kow CY, Ng JS, et al. Correlation of anterior interbody graft choice with patient-reported outcomes in cervical spine trauma. Global Spine Journal. 2019;9(7):735-742. doi:10.1177/2192568218828720

No

No

No

No

No

No

No

No
### Table 1. (continued)

| Article                                                                 | Study design          | Objective of study and grading/classification methods                                                                 | Presence and type of reliability analysis |
|------------------------------------------------------------------------|-----------------------|------------------------------------------------------------------------------------------------------------------------|------------------------------------------|
| Kawasaki Y, Seichi A, Zhang L, Tani S, Kimura A. Dynamic changes of cauda equina motion before and after decompressive laminectomy for lumbar spinal stenosis with redundant nerve roots: cauda equina activation sign. Global Spine Journal. 2019;9(6):619-623. doi:10.1177/2192568218821344 | Cross-sectional observational study | Radiographic evidence of postoperative interbody fusion and stability was assessed 6 months post-surgery independently by 2 investigators (HL and CK). The objective of the study was to define a criterion for achieving successful decompression of lumbar spinal stenosis (LSS) using intraoperative ultrasonography (IOUS) and to investigate the pathogenesis of redundant nerve roots (RNRs) based on the ultrasonographic findings.  - Cauda equina motion was stratified into one of 3 types  - The severity of spinal stenosis was assessed using grading system (A-D) proposed by Schizas et al. | No |
| Kanna RM, Shetty AP, Rajasekaran S. Predictors of successful outcomes of selective nerve root blocks for acute lumbar disc herniation. Global Spine Journal. 2019;9(5):473-479. doi:10.1177/219256821880050 | Prospective observational cohort | The authors describe a case series of patients with thoracolumbar fractures dislocations without spinal cord injury: classification and principles of management. Global Spine Journal. 2019;2192568219890568. doi:10.1177/2192568219890568 | No |
| Kawaguchi Y, Kitajima I, Nakano M, et al. Increase of the serum FGF-23 in ossification of the posterior longitudinal ligament. Global Spine Journal. 2019;9(5):492-498. doi:10.1177/2192568218801015 | Case-control | The objective of the study was to determine the factors that would predict successful outcomes in patients undergoing selective nerve root block (SNRB). Radiographic parameters included:  - Level of disc prolapse  - Position of disc in the axial T2 image  - Type of disc herniation  - Pfirrmann’s grade of the affected disc  - Presence of lumbosacral transitional vertebra at the level of the disc herniation Classification of nerve block radiculogram into 4 types:  - “Arm,” “arrow,” “linear,” and “splash” | No |
| Matsumoto T, Okuda S, Nagamoto Y, Sugiuara T, Takahashi Y, Iwasaki M. Effects of concomitant decompression adjacent to a posterior lumbar interbody fusion segment on clinical and radiologic outcomes: comparative analysis 5 years after surgery. Global Spine Journal. 2019;9(5):505-511. doi:10.1177/2192568218803324 | Retrospective cohort | The objective of the study was to examine the effects of concomitant decompression adjacent to the PLIF segment on the clinical and radiological outcomes 5 years after surgery.  - The extent of the concomitant posterior decompression at L3/4 was determined by preoperative myelography  - Disc height, vertebral slippage, range of motion (ROM) of the disc angle, posterior opening angle, segmental lordotic angle, the presence of the intradiscal vacuum phenomenon (IVP) at the L3/4 level, and lumbar lordosis at L1-S were measured on lateral radiograph (Figure 1A) | No |
| Oitment C, Kwok D, Steyn C. Calcified thoracic disc herniations in the elderly: revisiting the laminectomy for single level disease. Global Spine Journal. 2019;9(5):527-531. doi:10.1177/2192568218806274 | Retrospective case series | The authors examined a series of myelopathic elderly patients with calcified thoracic disc herniations who received a noninstrumented laminectomy without decompression of the disc space and hypothesize that patients who are medically unfit for extensive surgery or have significant medical comorbidities precluding them from direct decompression may benefit from a laminectomy without decompression of the disc space.  - Calcification was determined preoperatively via CT or intraoperatively  - ASIA scores were determined by reviewing clinical notes | No |
| Kanna RM, Raja DC, Shetty AP, Rajasekaran S. Thoracolumbar fracture dislocations without cord injury: classification and principles of management. Global Spine Journal. 2019;2192568219890568. doi:10.1177/2192568219890568 | Retrospective cohort | The authors describe a case series of patients with thoracolumbar fracture dislocation (TLFD) without cord injury and describe the different injury mechanisms, radiological subtypes, and principles of management.  - Thoracic and lumbar fractures were classified based on the direction of translation into 4 types: coronal translation (type 1), sagittal translation (type 2), combined translation—antero (type 3a), and combined translation—retro (type 3b) | No |
| Camino Willhuber G, Elizondo C, Slullitel P. Analysis of postoperative complications in spinal surgery, hospital length of stay, and unplanned readmission: application of Dindo-Clavien classification to spine surgery. Global Spine Journal. 2019;2192568219890568. doi:10.1177/2192568219890568 | Retrospective study | The authors sought to assess the 90-day postoperative complication rate following spinal surgeries by applying the Dindo-Clavien classification system (complication grade) as well as to analyze the unplanned readmission rate and total hospitalization LOS in patients with and without complications (Table 1) | No |

(continued)
Table 1. (continued)

| Article | Study design | Objective of study and grading/classification methods | Presence and type of reliability analysis |
|---------|--------------|-------------------------------------------------------|------------------------------------------|
| Spine Journal. 2019;9(3):279-286. doi:10.1177/2192568218792053 | Retrospective cohort | The authors sought to describe a novel facet-based motion segment classification system that significantly predicted postoperative segmental lordosis after lateral lumbar interbody fusion. Each facet was assigned a facet grade (min = 0, max = 2), and the sum of the bilateral facet grades was the final motion segment grade (MSG; min = 0, max = 4) (Table I). | No |
| Acosta FL, Mehta VA, Arakelyan A, et al. A novel lumbar motion segment classification to predict changes in segmental sagittal alignment after lateral interbody fixation. Global Spine Journal. 2017;7(7):642-647. doi:10.1177/2192568217773925 | Retrospective case-control | The objective of the study was to assess the clinical outcome of radiation therapy (RT) for spinal metastases in a retrospective case control study and to identify prognostic factors for pain control in patients with spinal metastases treated with RT. Pain grades were assessed using a drug grading level based on World Health Organization (WHO) standards. Drug grading using the WHO scale was as follows: 0, no drug; 1, nonsteroidal anti-inflammatory drugs; 2, weak opioid; 3, strong opioid (Table I). Authors used Tokuhashi's Evaluation System for the Prognosis of Metastatic Spine Tumor (Table 2). | No |
| Matsumura A, Hoshi M, Takami M, Tashiro T, Nakamura H. Radiation therapy without surgery for spinal metastases: clinical outcome and prognostic factors analysis for pain control. Global Spine Journal. 2012;2(3):137-141. doi:10.1055/s-0032-1326948 | Retrospective | The objective of the study was to investigate the clinical and radiographic outcomes to evaluate the efficacy of S2 alar screws in surgery for correction of adult spinal deformity (ASD). X-ray and CT reviewed by 2 radiologists to determine instrumentation and fusion-related radiographic outcomes (Tables 2 and 3). | No |
| Nakazawa T, Inoue G, Imura T, et al. Radiographic and clinical outcomes from the use of S2 alar screws in surgery for adult spinal deformity. Global Spine Journal. 2018;8(7):668-675. doi:10.1177/2192568218762378 | Retrospective observational | The authors sought to assess the placement accuracy and potentially influencing factors of 3-dimensionally navigated versus conventionally inserted pedicle screws. | No |
| Kraus M, Weiskopf J, Dreyhaupt J, Krischak G, Gebhard F. Computer-aided surgery does not increase the accuracy of dorsal pedicle screw placement in the thoracic and lumbar spine: a retrospective analysis of 2,003 pedicle screws in a level I trauma center. Global Spine Journal. 2015;5(2):093-101. doi:10.1055/s-0034-1396430 | Retrospective | The objective of the study was to evaluate the usefulness of Tokuhashi scoring (TS) system by comparing the predicted and real survival times and analyze the survival time according to the type of tumor (Table I). Preoperative TS was performed in all and patients were classified into 3 groups according to TS; group 1 (TS 0-8), group 2 (TS 9-11), and group 3 (TS 12-15). Imaging of the spine was assessed to define the number of metastases, location, and presence of cord or root compression. | No |
| Gruenberg M, Mereles ME, Willhuber GOC, Valacco M, Petracchi MG, Solá CA. Usefulness of Tokuhashi score in survival prediction of patients operated for vertebral metastatic disease. Global Spine Journal. 2017;7(3):260-265. doi:10.1177/2192568217762387 | Retrospective | The objective of the study was to develop a scoring system to estimate proximal junctional kyphosis (PJ) risk (Figure 2). | No |
| Lafage R, Beyer G, Schwab F, et al. Risk factor analysis for proximal junctional kyphosis after adult spinal deformity surgery: a new simple scoring system to identify high-risk patients. Global Spine Journal. 2019;21956821988235. doi:10.1177/2192568219882350 | Retrospective | The authors sought to describe the surgical outcomes in patients with high preoperative Spinal Instability Neoplastic Score (SINS) secondary to spinal giant cell tumors (GCT). Frankel classification was used to determine patients’ preoperative functional status, and the SINS was used to evaluate the degree of spinal instability. Tumor location, preoperative tumor volume, Enneking classification, and evidence of pathologic vertebral body fracture also obtained. | No |
| Elder B, Sankey E, Goodwin C, et al. Surgical outcomes in patients with high preoperative Spinal Instability Neoplastic Score (SINS) secondary to spinal giant cell tumors. Global Spine Journal. 2016;6(1):021-028. doi:10.1055/s-0035-1555657 | Retrospective | The objective of the study was to examine the usefulness of frailty scoring in elderly patients undergoing surgery for spinal tuberculosis. | No |

(continued)
Science (paid directly to institution), Motion Metrix (paid directly to institution); North American Spine Society: committee member; Lumbar Spine Society: Co-chair Research committee, AOSpine Knowledge Forum Degenerative: Associate member; AOSNA Research committee- committee member.

**Funding**

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

**ORCID iD**

Zorica Buser, PhD [https://orcid.org/0000-0002-5680-0643](https://orcid.org/0000-0002-5680-0643)

**Tables**

| Article | Study design | Objective of study and grading/classification methods | Presence and type of reliability analysis |
|---------|--------------|-----------------------------------------------------|-----------------------------------------|
| Sciubba D, Jain A, Kebaish KM, et al. Development of a preoperative adult spinal deformity comorbidity score that correlates with common quality and value metrics: length of stay, major complications, and patient-reported outcomes. *Global Spine J.* 2019;219256821989495. doi:10.1177/219256821989495 | Retrospective | • Modified frailty score (MFS) was calculated for each patient and was used to predict 30 day mortality (Table 1).  
• Other parameters recorded were American Society of Anesthesiologists (ASA) grade, operative time, blood loss, length of hospital stay, intensive care unit (ICU) stay and preoperative neurological deficit using Frankel system (Table 3).  
The authors developed a novel comorbidity score specific to patients with adult spinal deformity (ASD) based on their preoperative health state and investigated whether it was associated with major complications, length of hospital stay, and self-reported outcomes after ASD surgery (Table 4). | No |
| Kato M, Namikawa T, Matsumura A, Konishi S, Nakamura H. Radiographic risk factors of reoperation following minimally invasive decompression for lumbar canal stenosis associated with degenerative scoliosis and spondylolisthesis. *Global Spine J.* 2017;7(6):498-505. doi:10.1177/2192568217699192 | Prospective cohort study | The objective of the study was to evaluate the appropriateness of microsurgical bilateral decompression via a unilateral approach (MBDU) for degenerative lumbar scoliosis (DLS) or degenerative spondylolisthesis (DS).  
• Clinical indications for MBDU were leg pain and/or leg numbness inducing intermittent claudication  
• The radiological indications were lumbar spinal stenosis, DLS with a Cobb’s angle $\geq 10^\circ$ or $\leq 20^\circ$, and DS with Meyerding grade $\leq 1$ and posterior opening $\leq 5^\circ$ during anterior flexion of the affected intervertebral level  
• Patients were divided into 3 groups according to their diagnosis | No |

**References**

1. Dettori JR, Norvell DC. Kappa and beyond: is there agreement? *Global Spine J.* 2020;10:499-501. doi:10.1177/2192568220911648
2. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol.* 2008;61(pt 1):29-48. doi:10.1348/000711006X126600
3. Abedi A, Mokkink LB, Zadegan SA, et al. Reliability and validity of the AOSpine Thoracolumbar Injury Classification System: a systematic review. *Global Spine J.* 2019;9:231-242. doi:10.1177/2192568218806847
4. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 1968;70: 213-220. doi:10.1037/h0026256
5. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539-549. doi:10.1007/s11136-010-9606-8
6. Abedi A, Prinsen CAC, Shah I, Buser Z, Wang JC. Performance properties of health-related measurement instruments in whiplash: systematic review protocol. *Syst Rev.* 2019;8:199. doi:10.1186/s13643-019-1119-0
7. Gwet KL. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters.* 4 ed. Advanced Analytics, LLC; 2014.