Global Spine Journal
2025, Vol. 0(0) 1-11
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21925682251384662
journals.sagepub.com/home/gsj

Radiographic Assessment of Successful Lumbar Spinal Fusion: A Systematic Review of Fusion Criteria in Randomized Trials

Alexander Yu, BS[1], Justin Tiao, BS[1], Charlene W. Cai, BS[1], Jonathan J. Huang, AB[1], Kareem Mohamed, BS[1], Ryan Hoang, BS[1], James Hong, BS[1], Daniel Berman, MD[1], Joshua Lee, MD[1], Luca Ambrosio, MD[2], Zorica Buser, PhD, MBA[3,4], Juan P. Cabrera, MD[5], Xiaolong Chen, MD, PhD[6], Chiara Cini, MD[7], Stipe Ćorluka, MD[8], Andreas K. Demetriades, MB, BChir[9], Ashish Diwan, PhD, FRACS[10,11], Amit Jain, MD, MBA[12], Jin-Sung Kim, MD, PhD[13], Xudong Li, MD, PhD[14], Sathish Muthu, MS, PhD[15,16,17], Javad Tavakoli, PhD[18], Gianluca Vadalà, MD, PhD[2], Patrick C. Hsieh, MD, MBA[19], Samuel K. Cho, MD[1], and AO Spine Knowledge Forum Degenerative

Abstract

Study Type: Systematic Scoping Review.

Introduction: Radiographic assessment is crucial for diagnosing symptomatic pseudarthrosis and evaluating spinal fusion outcomes, yet no consensus exists on defining successful lumbar fusion. This scoping review presents criteria for imaging-based assessment after posterolateral and interbody lumbar fusion, aiming to guide consistent evaluation methods.

[1] Department of Orthopaedic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2] Operative Research Unit of Orthopaedic and Trauma Surgery, Fondazione Policlinico Universitario Campus Bio-Medico, Rome, Italy
[3] Gerling Institute, Brooklyn, NY, USA
[4] Department of Orthopedic Surgery, NYU Grossman School of Medicine, New York, NY, USA
[5] Department of Neurosurgery, Hospital Clínico Regional de Concepción, Concepción, Chile
[6] Department of Orthopedics, Xuanwu Hospital Capital Medical University, Beijing, China
[7] Department of Spine Surgery, IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy
[8] Spinal Surgery Division, Department of Traumatology, University Hospital Centre Sestre Milosrdnice, Zagreb, Croatia
[9] Department of Neurosurgery, Royal Infirmary of Edinburgh, Edinburgh, UK
[10] Department of Orthopaedic Surgery, Spine Service, St. George & Sutherland Clinical School, University of New South Wales Australia, Sydney, NSW, Australia
[11] Department of Orthopaedic Surgery, University of Adelaide & the Royal Adelaide Hospital, Adelaide, SA, Australia
[12] Department of Orthopaedic Surgery, Johns Hopkins University, Baltimore, MD, USA
[13] Department of Neurosurgery, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul, Republic of Korea
[14] Department of Orthopaedic Surgery, University of Virginia, Charlottesville, VA, USA
[15] Department of Orthopaedics, Government Medical College, Karur, India
[16] Department of Biotechnology, Karpagam Academy of Higher Education, Coimbatore, India
[17] Department of Spine Surgery, Orthopaedic Research Group, Coimbatore, India
[18] Department of Biomedical Engineering, School of Engineering, RMIT University, Melbourne, VIC, Australia
[19] Department of Neurological Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Corresponding Author:
Samuel K. Cho, MD, Department of Orthopaedic Surgery, Icahn School of Medicine at Mount Sinai, 787 11th Avenue, 7th Floor New York, NY 10019, USA.
Email: samuel.cho@mountsinai.org

Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 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).

Methods: Following PRISMA guidelines, a comprehensive search of Medline, Embase, and Scopus identified eligible randomized controlled trials and Federal Drug Administration clinical trials involving lumbar fusion. Studies on revision surgeries, non-lumbar fusions, adult spinal deformity, traumatic fractures, tumors, infections, or lacking defined fusion assessment methods were excluded. Data extraction focused on classification and descriptive systems of fusion evaluation, analyzing parameters such as bony bridging, angular motion, translation, hardware failure, cage migration, radiolucency, and cleft within the fusion mass.
Results: A total of 142 articles (1995-2024) were reviewed. Computerized tomography was the most common imaging modality (102, 71.8%), followed by static (96, 67.6%) and dynamic radiographs (88, 62%). Descriptive criteria were used in 108 studies (76.1%) and classification systems in 47 (33.1%). Interbody fusion was assessed in 90 articles (63.4%) and posterolateral fusion in 68 (47.9%). Bony bridging continuity was the most reported descriptive criterion (105, 73.9%), followed by angular motion (72, 50.7%) and translation (43, 30.3%). Radiolucency was reported around the cage (31, 21.8%), pedicle screws (17, 11.9%), and within the fusion mass (36, 25.4%) to describe nonunion. Common classification systems included Bridwell (10 studies), Brantigan (9), and Lenke (9).
Conclusions: This scoping review highlights the variability in lumbar fusion assessment across RCTs and FDA trials. Over time, assessment methods have evolved from static radiographs to greater use of dynamic imaging and classification systems in the mid-2000s, with CT emerging as the dominant modality in the past decade. Despite these advancements, fusion assessment criteria remain inconsistent across studies.

Keywords
lumbar fusion, classification, radiographic assessment, spine surgery, imaging

Introduction

Lumbar fusion surgery has become an increasingly common procedure for the treatment of various spinal pathologies, including degenerative disc disease, spondylolisthesis, and spinal stenosis, among others.[1-3] Its prevalence has risen significantly over the past three decades, driven by advancements in surgical techniques, instrumentation, imaging modalities, as well as implant and biologic materials.[4,5] A primary goal of lumbar fusion is to achieve solid osseous union between vertebral segments, thereby stabilizing the spine and alleviating pain or neurological symptoms.[2,6] However, despite its widespread use, challenges remain in consistently defining and assessing successful fusion outcomes, particularly through radiographic assessment modalities.

Common indicators of successful fusion by radiographic assessment are the presence of continuous bony bridging between the fused vertebral segments, suggesting solid bone healing and stability[7,8]. Additionally, a lack of visual gaps at the fusion site and minimal or no movement between vertebrae during flexion-extension imaging are used to infer successful fusion.[9] Advancing imaging modalities, such as computed tomography (CT) scans, offer more detailed visualization of implant integration and bone growth, further allowing for a more precise evaluation of the fusion mass[7,10,11] Despite these tools, subjective interpretation and the absence of universally accepted assessment standards contribute to ongoing challenges in reliably defining fusion success.[12-14]

Variations in radiographic evaluation methods have resulted in significant methodological inconsistencies across studies, potentially limiting the comparability, relevance, and generalizability of their findings. Prior reviews have highlighted the heterogeneity in imaging techniques, criteria, and classifications used to define fusion. For example, a review of 374 studies identified over 250 combinations of criteria, emphasizing the absence of standardized definitions.[14] Another review of 187 studies similarly found frequent reliance on bony bridging, motion assessment, and radiolucency. Still, inconsistent use of classification systems such as Lenke and Christensen was noted.[13] These variations underscore the need for higher-quality evidence to establish consistent and reproducible definitions of fusion success.

The purpose of this study is to conduct a systematic scoping review of fusion assessment methods, focusing exclusively on higher-quality evidence, including randomized controlled trials (RCTs), and Federal Drug Administration (FDA) regulated clinical trials. By analyzing how successful fusion is evaluated in high-level evidence, this review aims to provide insights into current practices and highlight methodological strengths and gaps. Ultimately, these findings seek to inform future efforts to establish standardized guidelines for post-operative lumbar fusion assessment, ensuring consistency and reliability in defining fusion success.

Methods

Search Strategy

On July 12, 2024, a comprehensive literature search of Medline, Embase, and Scopus for peer-reviewed journal articles was performed. Peer-reviewed original articles were queried with deduplication performed automatically. The search strategy was adapted from systematic reviews by Duits et al and Lehr et al. including keywords, synonyms, and

variations of the terms: "spine", "lumbar", "fusion", "arthrodesis", "posterolateral", and "interbody".[13,14]

The following PICOT acronym was used:

P (Population): Adult patients receiving lumbar fusion for degenerative pathologies.

I (Intervention): Posterolateral lumbar fusion and/or interbody fusion surgery.

C (Comparison): Fusion vs non-fusion.

O (Outcome): Classification and descriptive systems of fusion evaluation, analyzing parameters

Such as bony bridging, angular motion, translation, hardware failure, cage migration, radiolucency, and cleft within the fusion mass.

T (Time): January 1995 to July 2024.

This literature review was reported in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines.[15] The PRISMA screening process is outlined in Figure 1.

Study Eligibility

Studies met inclusion criteria if they utilized lumbar fusion interventions such as posterolateral fusion and/or interbody fusion in adult populations with lumbar spine degenerative disease. RCTs, prospective randomized studies, and FDA clinical trials involving an Investigational Device Exemption (IDE) were selected for inclusion as these categories provide a superior quality of clinical evidence, providing a more

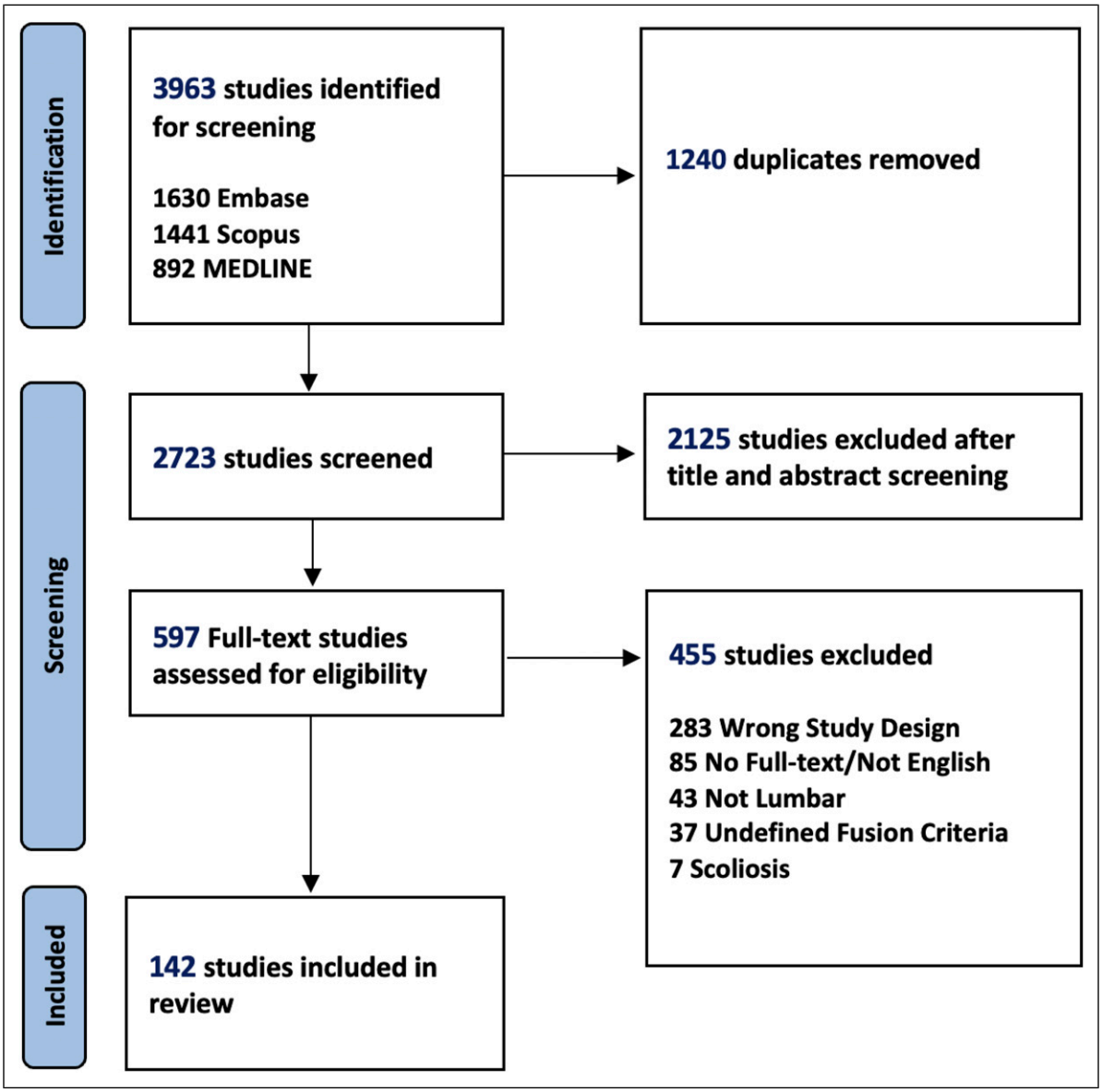


Figure 1. PRISMA Flow Chart of the Study Selection Process

accurate reflection of fusion assessment. Studies were included only if they were written in English and if full-text was available for review. Cases involving revision surgery were excluded, as were those focused on cervical or thoracic fusion. Studies were also ineligible if diagnoses involved scoliosis, traumatic fractures, or pathological conditions such as tumor growth or infection (e.g. tuberculosis). Studies with no definition of fusion assessment methods were also excluded.

Processing of Studies and Data Extraction

Two independent reviewers screened each study based on titles, abstracts, and full texts according to the inclusion and exclusion criteria. Any conflicting decisions between two reviewers were resolved through discussion with a third reviewer. Reasons for ineligibility or exclusion of studies were documented. The senior author (S.K.C) confirmed the included studies.

Data extraction was independently performed by the reviewers using standardized data extraction forms. Extracted data included methods used to define fusion based on two systems: classification and descriptive. The classification system utilizes an established grading or scoring system to assess the quality of fusion (Table 1). The descriptive system involves a practical description of fusion that is not addressed by any classification system. Subcategories within the descriptive system were developed by a spine fellow and the senior author, a board-certified spine surgeon, to ensure comprehensiveness and reflect current methodologies for evaluating various key indicators of fusion quality (Table 2).

Risk of Bias Assessment

The methodological quality of each included study was independently assessed by two reviewers using the Cochrane Risk of Bias Tool (RoB 2.0) for randomized controlled trials. For non-randomized FDA trials, the ROBINS-I tool was applied. Discrepancies were resolved by consensus or third-party adjudication. Each study was rated across domains including bias due to the randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of reported result. A summary judgment of "low," "some concerns," or "high" risk of bias was assigned to each study and can be found in the supplemental material. Importantly, our review focuses exclusively on the methodological quality of the studies and does not assess clinical or efficacy outcomes.

Table 1. Criteria of Common Classification Systems

Name	Year Introduced	Imaging Modality	Fusion Type	Criteria
Christensen	1988	Radiograph	Posterolateral	Complete fusion: Presence of continuous bone bridging across the fusion site, indicating a successful fusion Questionable fusion: Some evidence of bone formation, but it is not sufficient to definitively confirm a complete fusion Definitive pseudarthrosis: Clear lack of bone bridging, indicating a failure of the fusion process
Lenke	1992	Radiograph	Posterolateral/ Interbody	A: definitely solid (solid big trabeculated bilateral fusion masses) B: possibly solid (unilateral large fusion mass with contralateral small fusion mass) C: probably not solid (bilateral small, thin fusion masses with apparent crack) D: definitely not solid (bilateral graft resorption or fusion mass with obvious bilateral pseudarthrosis)
Bridwell	1993	Radiograph	Posterolateral/ Interbody	Grade I: Fused with remodeling and trabeculae present Grade II: Graft intact, not fully remodeled and incorporated, but no lucency Grade III: Graft intact, potential lucency at the top or bottom of the graft Grade IV: Fusion absent with collapse/resorption of the graft
Brantigan	1994	CT	Interbody	Grade 1: Definite solid fusion with no radiolucent lines and no motion Grade 2: Probable solid fusion with some radiolucent lines but no motion Grade 3: Possible pseudarthrosis with radiolucent lines and some motion Grade 4: Definite pseudarthrosis with clear radiolucent lines and significant motion
Glassman	2007	CT/ Radiograph	Posterolateral	Solid bilateral fusion: Complete fusion on both sides of the spine Solid unilateral fusion: Complete fusion on one side of the spine Partial bilateral fusion: Incomplete fusion on both sides of the spine Partial unilateral fusion: Incomplete fusion on one side of the spine Non-fusion: No evidence of fusion on either side of the spine

Table 2. Criteria of Descriptive System Subcategories

Indicator	Criteria
Continuity of bony bridging	Formation of new bone spanning the operated disc space, connecting adjacent vertebrae, indicating solid fusion
Angular motion	Degree of intervertebral motion between the fused segments on flexion-extension radiographs
Translation	Translational motion between vertebral bodies observed on flexion-extension radiographs
Hardware failure	Evaluation of screws, rods, or plates placed during the fusion process that may become loose or break, indicating instability and bony non-union
Subsidence	Reduction in the height of the operated intervertebral disc space, often due to compression of the graft or cage into the vertebral endplates
Cage migration	Complication where the implanted cage moves out of its intended position within the intervertebral space, indicating a failure to fuse
Radiolucency around cage	Radiolucent zone around the cage on radiographs, suggesting a lack of bone integration, associated with an increased risk of pseudarthrosis
Radiolucency around screws	Radiolucent zone around screws on radiographs, suggesting screw loosening, associated with an increased risk of pseudarthrosis
Cleft in fusion Mass	Assessed with CT imaging, indicating a gap or cleft within the fusion mass, associated with incomplete bone bridging and potential pseudarthrosis

Results

Search Results

After removing duplicates, the initial database searches yielded 3963 studies. Following the title and abstract screening, 2125 studies were excluded. The remaining 597 full-text reports were assessed, and 142 studies were deemed suitable for inclusion in the final analysis. Figure 1 displays a PRISMA flowchart detailing the screening process and literature search outcomes.

Study Characteristics

Early studies dated between 1995 and 2004, accounted for 24 trials (16.9%) (Table 3). This number rose to 54 trials (38%) between 2005-2014 and 64 trials (45.1%) between 2015-2024 (Table 3). All 142 studies used static radiographs, dynamic radiographs, and/or CT as their imaging modalities for assessing fusion. CT was the most frequently reported modality (n = 102, 71.8%), followed by static radiographs (n = 96, 67.6%) and dynamic radiographs (n = 88, 62%). The use of CT increased over time, appearing in 33.3% of studies from 1995-2004. This rose to 74.1% in 2005-2014 and further to 84.4% in 2015-2024. Among the fusion assessment methods, descriptive criteria were the most frequent method for defining fusion classification (n = 108, 76.1%). In comparison, 47 articles used a commonly established classification system (33.1%). Sixteen studies employed both descriptive criteria and a classification system. Interbody fusion was the most common approach with 90 articles (63.4%), followed by 68 articles describing the posterolateral approach (47.9%).

Eight descriptive criteria (Figure 2) for assessing fusion were identified and are summarized in Table 4. Indicators of fusion quality such as continuity of bony bridging and radiolucency were typically qualitative. Continuity of bony bridging was the most frequently reported criterion, appearing in 105 articles (73.9%). Dynamic instability criteria, including angular motion and translation, were quantitative but showed variability in cutoff values. Angular motion thresholds ranged from <2° to <5° (72 articles, 50.7%). Translation cutoffs varied between 1 and 3 mm and were used in 43 articles (30.3%). The static instability criteria, consisting of hardware

Table 3. Study Characteristics

	No. (%) of articles
Total #	142
Type of fusion	
Posterolateral	68 (47.9%)
Interbody	90 (63.4%)
Imaging modality	
Static radiographs	96 (67.6%)
Dynamic radiographs	88 (62.0%)
CT	102 (71.8%)
Fusion assessment method	
Descriptive criteria	108 (76.1%)
Classification system	47 (33.1%)
Decade of publication	
1995 - 2004	24 (16.9%)
2005 - 2014	54 (38.0%)
2015 - 2024	64 (45.1%)
Country of origin	
USA	33
China	27
South Korea	14
Japan	12
Denmark	10
Others	45

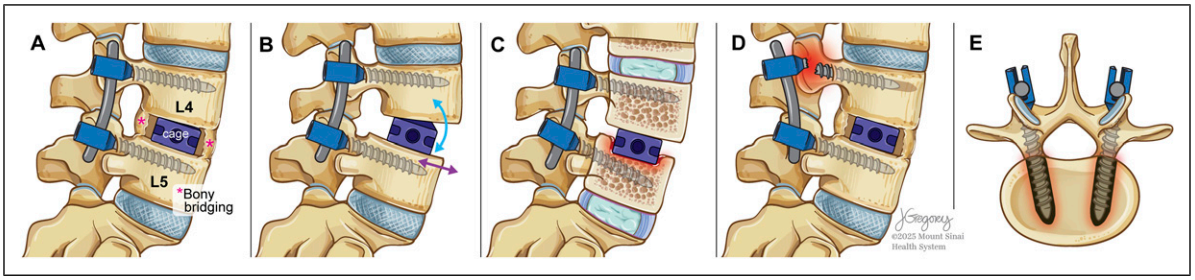


Figure 2. Representative Radiographic Features Commonly Used to Assess Lumbar Spinal Fusion. (A) Continuity of Bony Bridging Between Vertebral Bodies, Indicating Osseous Union (B) Angular Motion and Translation on Dynamic Radiographs, Used to Evaluate Spinal Stability (C) Evidence of Cage or Graft Subsidence Into Adjacent Endplates (D) Hardware Failure, Such as Pedicle Screw Loosening or Breakage (E) Radiolucency Around Screws, Suggesting Potential Nonunion or Loosening

failure and subsidence, were reported less frequently, appearing in 17.6% (n = 25) and 20.4% (n = 29) of studies, respectively. Radiolucency around the cage, around pedicle screws, and within the fusion mass were used in 31 (21.8%), 17 (11.9%), and 36 (25.4%) studies, respectively.

Fifteen distinct classification systems were identified, each utilizing specific criteria to define fusion, typically consisting of 3- to 5-point grading scales (Table 1 and Table 5). The Bridwell and Brantigan classifications were among the most frequently employed systems. The Bridwell classification, reported in 10 articles, categorized anterior fusion into four grades (I-IV), emphasizing features such as remodeling, trabeculation, and the presence or absence of lucency.[16] The Brantigan classification, utilized in nine articles, relied on radiographic characteristics like bridging bone and the absence of radiolucency, with fusion grades ranging from definite solid union to clear pseudarthrosis.[17] Other systems included the Christensen classification, which was applied in six articles and focused on the presence of continuous bone bridging to define complete fusion.[18] The Lenke classification, used in nine articles, categorized posterior fusion based on the size and quality of trabeculated fusion masses.[19] The Glassman classification, cited in seven articles, assessed fusion bilaterally and unilaterally, ranging from solid bilateral fusion to non-fusion.[19] Ten articles in the miscellaneous category included systems used in only a single study. All classification systems tended to emphasize positive indicators of union, such as bone density and maturation while offering structured frameworks for evaluating fusion outcomes.

Temporal analysis revealed distinct shifts in fusion assessment practices over the three-decade study period. Use of descriptive criteria remained consistently high across all time intervals, with continuity of bony bridging being the most frequently reported criterion in each decade (22 studies from 1995-2004, 42 from 2005-2014, and 41 from 2015-2024). Meanwhile, indicators such as angular motion, translation, and subsidence saw modest increases over time, reflecting growing interest in dynamic and structural parameters. Notably, reliance on classification systems increased significantly in recent years. While classifications such as Brantigan and Bridwell were absent or rare in studies prior to 2005, their use rose sharply after 2014—Bridwell was cited in 9 studies and Brantigan in 9, all in the most recent decade. Similarly, Lenke and Glassman classifications showed increased use.

Table 4. Descriptive Criteria Utilized for Assessing Fusion

	No. of articles			No. (%) of articles
Categorical Criteria	1995 - 2004	2005 - 2014	2015 - 2024	Overall
Continuity of bony bridging	22	42	41	105 (73.9%)
Dynamic instability				
Angular motion	13	33	26	72 (50.7%)
Translation	9	17	17	43 (30.3%)
Static instability				
Hardware failure	3	11	11	25 (17.6%)
Subsidence	2	9	18	29 (20.4%)
Radiolucency				
Around the cage	4	12	15	31 (21.8%)
Pedicle screws	5	5	7	17 (11.9%)
Cleft in fusion Mass	8	14	14	36 (25.4%)

Table 5. Classifications Utilized for Assessing Fusion

	No. of articles			
Classifications	1995 - 2004	2005 - 2014	2015 - 2024	Overall
Brantigan	0	0	9	9
Bridwell	0	1	9	10
Christensen	3	3	0	6
Glassman	0	3	4	7
Lenke	1	2	6	9
Miscellaneous	0	4	6	10

Discussion

Evaluating the status of bony fusion following spinal fusion surgery is essential. While open surgical exploration is the most definitive technique, noninvasive diagnostic imaging is widely used for routine assessment. However, there is no universally accepted definition of successful fusion in this context. The various imaging modalities come with distinct strengths and weaknesses, resulting in reported fusion rates that depend on varied criteria supported by limited clinical evidence. To explore the postoperative evaluation of lumbar spinal fusion, this study examined how the highest levels of evidence assessed successful postoperative fusion by conducting a scoping review of 142 RCTs and FDA trials published from 1995-2024.

The results demonstrate that interest in fusion has grown, with 45.1% of studies published between 2015 and 2024, consistent with the 95% rise in lumbar fusion procedures billed to Medicare between 2000 and 2019.[20] CT was the most used imaging modality (71.8%) due to clinical guidelines and its high sensitivity,[21] while static and dynamic radiographs (67.6% and 62.0%) may have been favored for their accessibility at follow-up.[22,23] Descriptive criteria (76.1%) predominated, while classification systems like Bridwell, Brantigan, and Lenke appeared in only a third of the studies, reflecting a lack of standardization. Continuity of bony bridging was the most reported descriptive criterion (73.9%) to define union, followed by lack of angular motion and translation as key indicators of spinal stability. Radiolucency, a marker of nonunion, was inconsistently noted, with a 25.4% prevalence in the fusion mass (i.e., cleft in fusion mass). These findings confirm the hypothesis of significant variability in assessing fusion status in the literature.

Prior meta-analyses regarding postoperative fusion methods in retrospective studies reported similar variances in the criteria used. Duits et al. (2024) found that 32% of interbody fusion studies used classification systems, primarily Bridwell and Brantigan, while 65% relied on descriptive criteria, with bony bridging being the most common.[14] Importantly, they revealed that descriptive criteria and classifications were applied in 256 unique combinations in the literature, with 45% of the combinations being used by only a single article. Similarly, Lehr et al (2022) found that 47% of posterolateral lumbar fusion studies used classification systems, mainly Lenke and Christensen, while 63% used descriptive criteria, again favoring bony bridging. Imaging usage included static radiographs (72%), dynamic radiographs (51%), and CT (35%).[13] As demonstrated in both the results and the prior literature, the variability in classification system usage and imaging modalities across studies highlights the need for more standardized criteria to improve comparability and reliability.

Our temporal analysis further revealed evolving patterns in both imaging modalities and fusion assessment strategies over the past three decades. Notably, the use of CT imaging increased substantially from 33.3% of studies in 1995-2004 to 84.4% in 2015-2024, reflecting growing reliance on higher-resolution, cross-sectional imaging to evaluate interbody fusion. In parallel, although descriptive criteria—particularly continuity of bony bridging—have remained the predominant method across all time periods, the use of formal classification systems rose sharply in the most recent decade, with Bridwell, Brantigan, and Lenke classifications increasingly adopted. This trend may reflect the maturing landscape of lumbar fusion research and a growing recognition of the need for reproducible grading tools amid technological advancements in imaging. However, despite these encouraging shifts, the continued dominance of non-standardized descriptive criteria underscores the field's lack of consensus.

Clinically, it is crucial to establish a standardized methodology for defining spinal fusion as many studies investigate the impact of various factors on fusion rates[13,14,24-28] The lack of uniform criteria across the literature complicates comparisons between studies. While previous efforts to standardize the assessment of bony fusion have been made, they predominantly focus on imaging modalities.[21,29-31] For instance, although CT scans are the most utilized imaging methodology for assessing fusion rates, variability in their accuracy persists.[14] This variability stems from several factors, including differences in imaging protocols, interobserver variability in interpretation, and the influence of metal artifacts from spinal implants, which can obscure bony details and lead to inconsistent assessments of fusion status.[32-34] Additionally, the sensitivity of CT scans in detecting pseudarthrosis may be limited in certain cases, particularly when subtle motion or radiolucency is present but not clearly visualized due to technical or anatomical constraints.[31] Advances in CT technology, such as helical scanning, multiplanar reconstructions, higher resolution, and reduced artifacts from metal implants, may contribute to further inconsistencies in the literature.[35-37] For example, newer techniques like dual-energy CT and iterative metal artifact reduction algorithms have shown promise in reducing implant-related artifacts, but their adoption and interpretation vary widely across institutions, leading to discrepancies in fusion assessment.[38-40] Although MRI studies on spinal fusion remain limited, MRI has been explored as an alternative imaging modality. MRI does not involve radiation exposure and provides detailed visualization

of neural and adjacent soft tissue structures. Some studies have examined its potential role in assessing fusion, with varying degrees of agreement compared to CT. Kitchen et al. reported moderate concordance between MRI and CT in evaluating fusion within interbody cages, particularly in coronal planes ($\kappa = 0.58$).[41] Similarly, Kröner et al. found that coronal MRI could visualize bony bridging through cages with high interobserver agreement (88%).[42] While MRI may address some limitations associated with CT, its use for fusion assessment is not yet widely established.

Fusion is a costly procedure, with the industry heavily invested in implants and biologics as surgeons pursue the "ideal fusion."[43-45] However, the absence of both consistency and consensus in evaluation methods complicates efforts to optimize outcomes and manage costs. These challenges underscore the need for standardized assessment criteria that integrate imaging modalities with clear guidelines for interpretation and reporting, ensuring consistency and reliability across studies and clinical practice.

The FDA has proposed its own criteria for defining fusion in its guidance document for manufacturers preparing an IDE for spinal systems.[46] This criteria was first established in 1998, with subsequent updates in 2000 and 2004.[47] According to the FDA, successful fusion is defined as: (1) evidence of bridging trabecular bone between the involved motion segments, (2) translational motion less than 3 mm, and (3) angular motion less than 5° demonstrated on X-ray (anterior-posterior, lateral, flexion, and extension views). For manufacturers utilizing an alternative radiographic modality, the FDA requires a demonstration of the validity and reliability of the chosen modality before its utilization as a primary study endpoint.[48] However, the current study found that only 30% of studies utilized translational motion as one of its criteria for assessing fusion, indicating that many studies do not adhere to FDA-recommended imaging criteria. Additionally, more than 70% of studies utilized CT scans to assess fusion, further underscoring the deviation from FDA recommendations. This suggests that the FDA guidelines may no longer reflect current clinical practices regarding the optimal imaging modality, as our findings indicate a significant rise in CT utilization from 2005 to 2024, well after these guidelines were established. These findings emphasize the importance of developing and validating standardized criteria for fusion assessment to ensure consistency across studies and clinical practices, ultimately enhancing the reliability and generalizability of research findings. Until a universally accepted definition is established, wider adoption of the FDA's recommendations with current imaging best practices could help standardize fusion assessment and enhance comparability across studies.

Another factor contributing to the variability in assessing fusion is the wide range of classification systems utilized. The investigation identified 15 unique classification systems. This aligns with previous literature, which has also reported varied classification systems used to assess fusion.[13,14] Among these, the Bridwell classification was the most commonly utilized, whereas other studies, such as Lehr et al. (2002), identified the Lenke classification to be the most prevalent.[13] Despite minor differences, most classification systems share common elements, such as evaluating bony continuity and the presence or absence of motion. However, subtle differences between these systems make it challenging to compare studies utilizing different criteria. Moreover, even within the same classification system, discrepancies in the imaging modalities used (e.g., CT scan vs X-ray) further limit comparability.[14]

To achieve true generalizability in studies assessing spinal fusion, a standardized and validated set of criteria, including imaging modalities, must be established. With rapid advancements in technology, minimizing human error in fusion assessment is becoming increasingly feasible. Several approaches warrant investigation to develop a comprehensive and standardized assessment framework. Standardized imaging protocols can enhance consistency by ensuring uniform imaging techniques, patient positioning, and interpretation across studies. Additionally, advanced image processing and AI-driven analysis offer objective and reproducible measures of fusion, reducing observer bias and improving accuracy.[49,50] Machine learning-based predictive modeling can further integrate diverse clinical and imaging factors to forecast fusion outcomes, facilitating a more standardized evaluation across different fusion types and patient populations.[51] Separately defining fusion subtypes, such as posterolateral or interbody fusion, is also essential, as it enables tailored assessment methods that improve specificity and generalizability. Furthermore, consolidating existing classification systems into a unified framework could enhance consistency and comparability in fusion research. Collectively, these strategies aim to refine and unify fusion assessment criteria, ultimately improving reliability and standardization in both clinical and research settings.

Limitations and Future Directions

This scoping review, while comprehensive, entails inherent limitations to be considered when interpreting the findings. First, the scope of this review is descriptive and classification-focused, rather than quantitative or accuracy-focused. This approach limits this study's ability to directly compare the effectiveness or precision of different lumbar fusion assessment methodologies. As such, the conclusions drawn are primarily based on the presence and descriptions of methodologies rather than their quantitative validation against clinical outcomes.

Moreover, the review was restricted to articles published in English, introducing a potential language bias. Additionally, the inclusion of older studies introduces another layer of complexity given that imaging technologies have significantly evolved over the decades covered by this review (1997-2024). Earlier studies utilized imaging modalities that, by today's standards, might be considered less precise. This technological discrepancy could impact the interpretation of certain fusion

assessment methods, as older imaging techniques may not provide the necessary level of detail required by modern descriptive criteria.

Another challenge is that not all instances of delayed fusion or radiographic pseudoarthrosis are clinically relevant.[52] Some patients with radiographic evidence of pseudoarthrosis remain asymptomatic, whereas others may require intervention. This highlights the need for an integrated approach to distinguish between unfused patients who require treatment and those with asymptomatic pseudoarthrosis who may not. Additionally, differing techniques in the surgical approach to lumbar fusion (e.g., anterior, posterior, transforaminal) as well as variations in cage types were not considered in this study's analysis.

Future research should prioritize several key areas. First, quantitative validation studies are essential to objectively compare the effectiveness and precision of different lumbar fusion assessment methodologies. This would help establish standardized criteria that can be universally adopted in clinical settings. Alongside this, there is a pressing need for developing or validating unified classification systems for lumbar fusion specific to the fusion type (interbody/posterolateral) and assessment method (radiography/CT), with clear measurement cut-offs to define instability. A standardized system would facilitate consistent assessments across studies and clinical practices, enhancing the reliability and comparability of research findings.

Another critical research direction is the integration of clinical outcome measures in fusion assessments. Beyond radiographic evaluation, studies should investigate how well different assessment methodologies correlate with functional outcomes, pain relief, and the need for reoperation. Prospective studies assessing clinical indicators of successful fusion—such as patient-reported outcomes, biomechanical assessments, and adjacent segment degeneration—could provide a more holistic understanding of fusion success beyond imaging findings alone.

Conclusion

This systematic scoping review highlights the considerable variability in assessing lumbar fusion success in RCTs and FDA-regulated trials. The evolution of assessment methods and imaging modalities since 1995 reflects both technological advancements in spinal fusion research. Early studies primarily relied on static radiographs and descriptive criteria, often lacking standardized classification systems. In the mid-2000s, there was a notable increase in dynamic radiographic assessments, along with greater adoption of classification systems. The most recent decade saw the most significant shift, with CT becoming the predominant imaging modality. Despite these advancements, fusion assessment criteria remain inconsistent across studies (Supplemental Material).

Acknowledgments

Jill K Gregory, MFA, CMI - Certified Medical Illustrator. Associate Director of Scholarly Publishing and Visualization. Gustave L. and Janet W. Levy Library.

ORCID iDs

Alexander Yu https://orcid.org/0000-0002-7246-2269
Justin Tiao https://orcid.org/0000-0002-9414-7120
Kareem Mohamed https://orcid.org/0009-0003-9667-704X
Ryan Hoang https://orcid.org/0000-0002-1212-6183
Juan P. Cabrera https://orcid.org/0000-0003-4685-6106
Stipe Ćorluka https://orcid.org/0000-0002-5732-1150
Andreas K. Demetriades https://orcid.org/0000-0002-2004-9448
Amit Jain https://orcid.org/0000-0002-9983-3365
Jin-Sung Kim https://orcid.org/0000-0001-5086-0875
Sathish Muthu https://orcid.org/0000-0002-7143-4354
Javad Tavakoli https://orcid.org/0000-0002-0696-6530
Patrick C. Hsieh https://orcid.org/0000-0002-7206-4842
Samuel K. Cho https://orcid.org/0000-0001-7511-2486

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Disclosures

Samuel K. Cho, MD, FAAOS
AAOS: Board or committee member
American Orthopaedic Association: Board or committee member
AOSpine North America: Board or committee member
Cervical Spine Research Society: Board or committee member
Globus Medical: IP royalties and Fellowship support
North American Spine Society: Board or committee member
Scoliosis Research Society: Board or committee member
Stryker: Paid consultant
Cerapedics: Fellowship support.

Supplemental Material

Supplemental material is available online.

References

1. Reid PC, Morr S, Kaiser MG. State of the union: a review of lumbar fusion indications and techniques for degenerative spine disease. *J Neurosurg Spine*. 2019;31(1):1-14.
2. Mobbs RJ, Phan K, Malham G, Seex K, Rao PJ. Lumbar interbody fusion: techniques, indications and comparison of interbody fusion options including PLIF, TLIF, MI-TLIF, OLIF/ATP, LLIF and ALIF. *J Spine Surg*. 2015;1(1):2-18.
3. Steinberger J, Skovrlj B, Caridi JM, Cho SK. The top 100 classic papers in lumbar spine surgery. *Spine*. 2015;40(10):740-747.
4. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and regional variations in lumbar spine surgery: 1992-2003. *Spine*. 2006;31(23):2707-2714.
5. Ponkilainen VT, Huttunen TT, Neva MH, Pekkanen L, Repo JP, Mattila VM. National trends in lumbar spine decompression and

fusion surgery in Finland, 1997-2018. *Acta Orthop*. 2021;92(2): 199-203.
6. Zigler JE, Delamarter RB. Does 360° lumbar spinal fusion improve long-term clinical outcomes after failure of conservative treatment in patients with functionally disabling single-level degenerative lumbar disc disease? Results of 5-year follow-up in 75 postoperative patients. *Int J Spine Surg*. 2013;7:e1-e7.
7. Lansford T, Park DK, Wind JJ, et al. High lumbar spinal fusion rates using cellular bone allograft irrespective of surgical approach. *Int J Spine Surg*. 2024;18(4):355-364.
8. Kim DH, Hwang RW, Lee GH, et al. Potential significance of facet joint fusion or posteromedial fusion observed on CT imaging following attempted posterolateral or posterior interbody fusion. *Spine J*. 2020;20(3):337-343.
9. Ao S, Zheng W, Wu J, et al. Comparison of Preliminary clinical outcomes between percutaneous endoscopic and minimally invasive transforaminal lumbar interbody fusion for lumbar degenerative diseases in a tertiary hospital: is percutaneous endoscopic procedure superior to MIS-TLIF? A prospective cohort study. *Int J Surg*. 2020;76:136-143.
10. Williams AL, Gornet MF, Burkus JK. CT evaluation of lumbar interbody fusion: current concepts. *AJNR Am J Neuroradiol*. 2005;26(8):2057-2066.
11. Nakashima H, Yukawa Y, Ito K, et al. Extension CT scan: its suitability for assessing fusion after posterior lumbar interbody fusion. *Eur Spine J*. 2011;20(9):1496-1502.
12. Sanghvi P, Wiener J, Steinmetz MP. P247. Development of a unified and comprehensive definition of successful spinal fusion. *Spine J*. 2024;24(9):S185-S186.
13. Lehr AM, Duits AAA, Reijnders MRL, et al. Assessment of posterolateral lumbar fusion: a systematic review of imaging-based fusion criteria. *JBJS Rev*. 2022;10(10):e22.
14. Duits AAA, van Urk PR, Lehr AM, et al. Radiologic assessment of interbody fusion: a systematic review on the use, reliability, and accuracy of current fusion criteria. *JBJS Rev*. 2024;12(1):e23.
15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J*. 2021;372:n71.
16. Bridwell K, Lenke L, McEnery K, Baldus C, Blanke K. Anterior fresh frozen structural allografts in the thoracic and lumbar spine: do they work if combined with posterior fusion and instrumentation in adult patients with kyphosis or anterior column defects? *Spine*. 1995;20(12):1410-1418.
17. Brantigan JW, Steffee AD. A carbon fiber implant to aid interbody lumbar fusion. Two-year clinical results in the first 26 patients. *Spine*. 1993;18(14):2106-2107.
18. Christensen FB, Laursen M, Gelineck J, Eiskjaer SP, Thomsen K, Bünger CE. Interobserver and intraobserver agreement of radiograph interpretation with and without pedicle screw implants: the need for a detailed classification system in posterolateral spinal fusion. *Spine*. 1976;26:543-544.
19. Glassman SD, Dimar JR, Carreon LY, Campbell MJ, Puno RM, Johnson JR. Initial fusion rates with recombinant human bone morphogenetic protein-2/compression resistant matrix and a hydroxyapatite and tricalcium phosphate/collagen carrier in posterolateral spinal fusion. *Spine*. 1976;30:1694-1698.
20. Singh R, Moore ML, Hallak H, et al. Recent trends in medicare utilization and reimbursement for lumbar fusion procedures: 2000-2019. *World Neurosurg*. 2022;165:e191-e196.
21. Choudhri TF, Mummaneni PV, Dhall SS, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 4: radiographic assessment of fusion status. *J Neurosurg Spine*. 2014;21(1):23-30.
22. Rutherford EE, Tarplett LJ, Davies EM, Harley JM, King LJ. Lumbar spine fusion and stabilization: hardware, techniques, and imaging appearances. *Radiographics*. 2007;27(6): 1737-1749.
23. Goodwin ML, Buchowski JM, Sciubba DM. Why X-rays? The importance of radiographs in spine surgery. *Spine J*. 2022; 22(11):1759-1767.
24. Pairojboriboon S, Niruthisard S, Chandhanayingyong C, Monsereenusorn C, Poopan S, Lo SFL. A comparison of transforaminal lumbar interbody fusion (TLIF) cage material on fusion rates: a systematic review and network meta-analysis. *World Neurosurg X*. 2024;23:100392.
25. Thomsen K, Christensen FB, Eiskjær SP, Hansen ES, Fruensgaard S, Bünger CE. Volvo award winner in clinical studies: the effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective, randomized clinical Study. *Spine*. 1997;22(24): 2813-2822.
26. Burkus JK, Transfeldt EE, Kitchel SH, Watkins RG, Balderston RA. Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic Protein-2. *Spine*. 2002;27(21):2396-2408.
27. Jiya T, Smit T, Deddens J, Mullender M. Posterior lumbar interbody fusion using nonresorbable poly-ether-ether-ketone versus resorbable Poly-l-Lactide-Co-d,l-Lactide fusion devices: a prospective, randomized Study to assess fusion and clinical outcome. *Spine*. 2009;34(3):233-237.
28. Kubota G, Kamoda H, Orita S, et al. Platelet-rich plasma enhances bone union in posterolateral lumbar fusion: a prospective randomized controlled trial. *Spine J*. 2019;19(2):e34-e40.
29. Gruskay JA, Webb ML, Grauer JN. Methods of evaluating lumbar and cervical fusion. *Spine J*. 2014;14(3):531-539.
30. Lee YP, Farhan SA, Musa A, Bhatia N. Pseudarthrosis in spine surgery: diagnosis and treatment. *Contemp Spine Surg*. 2019; 20(8):1-7.
31. Peters MJM, Bastiaenen CHG, Brans BT, Weijers RE, Willems PC. The diagnostic accuracy of imaging modalities to detect pseudarthrosis after spinal fusion-a systematic review and meta-analysis of the literature. *Skelet Radiol*. 2019;48(10):1499-1510.
32. Ho JM, Ben-Galim PJ, Weiner BK, et al. Toward the establishment of optimal computed tomographic parameters for the assessment of lumbar spinal fusion. *Spine J*. 2011;11(7):636-640.
33. Carreon LY, Djurasovic M, Glassman SD, Sailer P. Diagnostic accuracy and reliability of fine-cut CT scans with reconstructions to determine the status of an instrumented posterolateral fusion

with surgical exploration as reference standard. *Spine*. 2007; 32(8):892-895.
34. Dangelmaier J, Schwaiger BJ, Gersing AS, et al. Dual layer computed tomography: reduction of metal artefacts from posterior spinal fusion using virtual monoenergetic imaging. *Eur J Radiol*. 2018;105:195-203.
35. Rubin GD, Shiau MC, Schmidt AJ, et al. Computed tomographic angiography: historical perspective and new state-of-the-art using multi detector-row helical computed tomography. *J Comput Assist Tomogr*. 1999;23(Suppl 1):S83-S90.
36. Douek PC, Boccalini S, Oei EHG, et al. Clinical applications of photon-counting CT: a review of pioneer studies and a glimpse into the future. *Radiology*. 2023;309(1):e222432.
37. Katsura M, Sato J, Akahane M, Kunimatsu A, Abe O. Current and novel techniques for metal artifact reduction at CT: practical guide for radiologists. *Radiographics*. 2018;38(2):450-461.
38. Long Z, DeLone DR, Kotsenas AL, et al. Clinical assessment of metal artifact reduction methods in dual-energy CT examinations of instrumented spines. *AJR Am J Roentgenol*. 2019; 212(2):395-401.
39. Kotsenas AL, Michalak GJ, DeLone DR, et al. CT metal artifact reduction in the spine: can an iterative reconstruction technique improve visualization? *AJNR Am J Neuroradiol*. 2015;36(11): 2184-2190.
40. Wang F, Zhang Y, Xue H, et al. Combined use of iterative reconstruction and monochromatic imaging in spinal fusion CT images. *Acta Radiol*. 2017;58(1):62-69.
41. Kitchen D, Rao PJ, Zotti M, et al. Fusion assessment by MRI in comparison with CT in anterior lumbar interbody fusion: a prospective study. *Glob Spine J*. 2018;8(6):586-592.
42. Kröner AH, Eyb R, Lange A, Lomoschitz K, Mahdi T, Engel A. Magnetic resonance imaging evaluation of posterior lumbar interbody fusion. *Spine*. 2006;31(12):1365-1371.
43. Martin BI, Mirza SK, Spina N, Spiker WR, Lawrence B, Brodke DS. Trends in lumbar fusion procedure rates and associated hospital costs for degenerative spinal diseases in the United States, 2004 to 2015. *Spine*. 2019;44(5):369-376.
44. Ghogawala Z, Whitmore RG, Watters WC, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 3: assessment of economic outcome. *J Neurosurg Spine*. 2014;21(1):14-22.
45. Pahlavan S, Berven S, Bederman SS. Variation in costs of spinal implants in United States academic medical centers. *Spine*. 2016;41(6):515-521.
46. Devlin VJ, Jean R, Peat CR, et al. Summary of the FDA virtual public workshop on spinal device clinical review held on September 17, 2021. *Spine J*. 2022;22(9):1423-1433.
47. Center for Devices, Radiological Health. *Guidance for Industry and FDA Staff: Spinal System 510(k)s*. U.S. Food and Drug Administration; 2020. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-and-fda-staff-spinal-system-510ks. Accessed 1 February 2025.
48. Center for Devices, Radiological Health. *Guidance Document for the Preparation of IDEs for Spinal Systems - Guidance for Industry and/or FDA Staff*. U.S. Food and Drug Administration; 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-document-preparation-ides-spinal-systems-guidance-industry-andor-fda-staff. Accessed 21 December 2024.
49. Park S, Kim JK, Chang MC, Park JJ, Yang JJ, Lee GW. Assessment of fusion after anterior cervical discectomy and fusion using convolutional neural network algorithm. *Spine*. 2022; 47(23):1645-1650.
50. Zhou S, Yao H, Ma C, et al. Artificial intelligence X-ray measurement technology of anatomical parameters related to Lumbosacral stability. *Eur J Radiol*. 2022;146(110071):110071.
51. Yuan S, Chen R, Liu X, et al. Artificial intelligence automatic measurement technology of Lumbosacral radiographic parameters. *Front Bioeng Biotechnol*. 2024;12:1404058.
52. Dhall SS, Choudhri TF, Eck JC, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 5: correlation between radiographic outcome and function. *J Neurosurg Spine*. 2014;21(1):31-36.