

# UC Davis

## UC Davis Previously Published Works

### Title

How Reliable is the Assessment of Fusion Status Following ACDF Using Dynamic Flexion-Extension Radiographs?

### Permalink

<https://escholarship.org/uc/item/3hw7d7p5>

### Authors

Martin, Christopher

Yoon, Sangwook

Alluri, Ram

et al.

### Publication Date

2024-12-05

### DOI

10.1177/21925682241303107

Peer reviewed

# How Reliable is the Assessment of Fusion Status Following ACDF Using Dynamic Flexion-Extension Radiographs?

Global Spine Journal  
2024, Vol. 0(0) 1–8  
© The Author(s) 2024  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/21925682241303107  
[journals.sagepub.com/home/gsj](https://journals.sagepub.com/home/gsj)



Christopher T. Martin, MD<sup>1</sup>, Sangwook Tim Yoon, MD<sup>2</sup> , Ram Kiran Alluri, MD<sup>3</sup> , Edward C. Benzel, MD<sup>4</sup>, Chris M. Bono, MD<sup>5</sup>, Samuel K. Cho, MD<sup>6</sup> , Dean Chou, MD<sup>7</sup>, Xiaolong Chen, MD, PhD<sup>8</sup> , Jason P.Y. Cheung, MD<sup>9</sup>, Juan P. Cabrera, MD<sup>10,11</sup> , Stipe Ćorluka, MD<sup>12,13</sup> , Andreas K. Demetriades, MD<sup>14</sup> , Matthew F. Gary, MD<sup>2,15</sup>, Zoher Ghogawala, MD<sup>16</sup> , Wael Hamouda, MD<sup>17,18</sup> , Inbo Han, MD<sup>19</sup> , Dimitri Hauri, PhD<sup>20</sup>, Patrick C. Hsieh, MD<sup>21</sup> , Amit Jain, MD, MBA<sup>22</sup> , Jun S. Kim, MD<sup>7</sup>, Hai V. Le, MD<sup>23</sup> , Philip K. Louie, MD<sup>24</sup> , Zhuojing Luo, MD<sup>25</sup>, Hans-Jörg Meisel, MD, PhD<sup>26</sup> , Sathish Muthu, MS, PhD<sup>27,28,29</sup> , Dal-Sung Ryu, MD<sup>30</sup>, Charles A. Sansur, MD<sup>31</sup>, Andrew J. Schoenfeld, MD<sup>32</sup>, Laura Scaramuzzo, MD<sup>33</sup> , Gregory D. Schroeder, MD<sup>34</sup>, Shanmuganathan Rajasekaran, PhD<sup>35</sup> , Veranis Sotiris, MD<sup>36</sup>, Gianluca Vadalà, MD, PhD<sup>37,38</sup>, Pieter-Paul A. Vergroesen, MD, PhD, MBA<sup>39</sup>, Jeffrey C. Wang, MD<sup>3,21</sup>, Yabin Wu, PhD<sup>20</sup> , K. Daniel Riew, MD<sup>40,41</sup> , and AO Spine Knowledge Forum Degenerative

<sup>1</sup> Department of Orthopedic Surgery, University of Minnesota, Minneapolis, MN, USA

<sup>2</sup> Department of Orthopaedics, Emory University, Atlanta, GA, USA

<sup>3</sup> Department of Orthopaedic Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

<sup>4</sup> Department of Neurosurgery, Cleveland Clinic Foundation, Cleveland, OH, USA

<sup>5</sup> Department of Orthopaedic Surgery, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA

<sup>6</sup> Department of Orthopedic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>7</sup> Department of Neurosurgery, Columbia University Irving Medical Center, New York, NY, USA

<sup>8</sup> Department of Orthopedic Surgery, Xuanwu Hospital Capital Medical University, Beijing, China

<sup>9</sup> Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong SAR, China

<sup>10</sup> Department of Neurosurgery, Hospital Clínico Regional de Concepción, Concepción, Chile

<sup>11</sup> Faculty of Medicine, University of Concepción, Concepción, Chile

<sup>12</sup> Spinal Surgery Division, Department of Traumatology, University Hospital Centre Sestre Milosrdnice, Zagreb, Croatia

<sup>13</sup> Department of Anatomy and Physiology, University of Applied Health Sciences, Zagreb, Croatia

<sup>14</sup> Edinburgh Spinal Surgery Outcome Studies Group, Department of Neurosurgery, Royal Infirmary Edinburgh, Edinburgh, UK

<sup>15</sup> Department of Neurosurgery, Emory University, Atlanta, GA, USA

<sup>16</sup> Department of Neurosurgery, Lahey Hospital and Medical Center, Burlington, MA, USA

<sup>17</sup> Department of Neurosurgery, Kasr Alainy Faculty of Medicine, Research, and Teaching Hospitals, Cairo University, Cairo, Egypt

<sup>18</sup> Neurological & Spinal Surgery Service, Security Forces Hospital, Dammam, Saudi Arabia

<sup>19</sup> Department of Neurosurgery, CHA University School of Medicine, CHA Bundang Medical Center, Seongnam-si, Republic of Korea

<sup>20</sup> Clinical Science, AO Innovation Translation Center, AO Foundation, Davos, Switzerland

<sup>21</sup> Department of Neurological Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

<sup>22</sup> Department of Orthopaedic Surgery, Johns Hopkins University, Baltimore, MD, USA

<sup>23</sup> Department of Orthopaedic Surgery, University of California, Sacramento, CA, USA

<sup>24</sup> Center for Neurosciences and Spine, Department of Neurosurgery, Virginia Mason Medical Center, Seattle, WA, USA

## Corresponding Author:

Christopher T. Martin, MD, Department of Orthopedic Surgery, University of Minnesota, 2450 Riverside Avenue South Suite R200. Minneapolis, MN 55454, USA.

Email: [mart1865@umn.edu](mailto:mart1865@umn.edu)



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>).

## Abstract

**Study Design:** Reliability study.

**Objectives:** The radiographic diagnosis of non-union is not standardized. Prior authors have suggested using a cutoff of <1 mm interspinous process motion (ISPM) on flexion-extension radiographs, but the ability of practicing surgeons to make these measurements reliably is not clear.

**Methods:** 29 practicing spine surgeons measured ISPM on 19 levels of ACDF from 9 patients. Surgeons relied on these measurements to report on fusion status. Inter-observer correlation co-efficients (ICC), standard error (SEM) and the minimum detectable difference (MD) of these measurements were calculated. We screened for clerical errors by checking measurements more than one standard deviation from the group mean.

**Results:** The ICC for ISPM was .76 (.64; .88) with a SEM of 1 mm and a MD of 2.76 mm. Agreement on fusion status was moderate, with an ICC of .6 (.44; .76). After screening for and removing clerical errors, the ICC improved to .82 (.71; .91), SEM improved to .83 mm, and MD improved to 2.29 mm. Six reviewers had an ICC >.9. The ICC from these high performing reviewers was .94 (.9; .97), SEM was .45 mm, and MD was 1.26 mm.

**Conclusions:** The MD of 2.29 mm in our study group was not precise enough to support a cutoff of <1 mm ISPM as the sole measurement technique in screening for non-union after ACDF, and there was only moderate agreement amongst surgeons on fusion status based on dynamic radiographs. More stringent techniques are necessary to avoid mis-diagnosing non-union in clinical studies. Future studies should consider auditing measurements to identify clerical errors.

## Keywords

fusion, ACDF, interspinous process motion

## Introduction

Fusion rates after anterior cervical discectomy and fusion (ACDF) have historically been reported to be very high for one- and two-level cases, with rates ranging up to 90-100% in some series.<sup>1</sup> However, the methods employed to identify non-unions are quite variable across studies, and have included assessing for the presence of bridging bone across the operative levels on static images or CT scans, inter-spinous process motion (ISPM) on dynamic flexion-extension radiographs, and change in Cobb angle of the operative segment on dynamic flexion-extension radiographs.<sup>2-4</sup> Given the variability in assessment methods, it is difficult to compare fusion

rates and outcomes across studies. For example, one study listed a 100% fusion rate based on the presence of bridging bone on CT scans, but also reported that two of those patients had screw back-out at the operative levels, and the authors did not have any dynamic image assessment.<sup>5</sup> In contrast, studies that included an assessment of motion on dynamic flexion-extension radiographs have reported 12 month fusion rates as low as 64-65%,<sup>6,7</sup> and multiple authors have concluded that using a CT scan alone tends to under-estimate non-union incidence.<sup>2-4,6-8</sup> Furthermore, fusion assessment with CT is particularly challenging in cases with metallic interbody devices or with non-bone graft materials such as ceramics given the scatter created on the image. Such significant discrepancies

<sup>25</sup> Department of Orthopaedic Surgery, Xijing Hospital, Forth Military Medical University, Xi'an, China

<sup>26</sup> Department of Neurosurgery, BG Klinikum Bergmannstrost, Halle, Germany

<sup>27</sup> Orthopaedic Research Group, Coimbatore, India

<sup>28</sup> Department of Orthopaedics, Government Medical College, Karur, India

<sup>29</sup> Department of Biotechnology, Karpagam Academy of Higher Education, Coimbatore, India

<sup>30</sup> Department of Neurosurgery, College of Medicine, Inha University, Incheon, Korea

<sup>31</sup> Department of Neurosurgery, University of Maryland School of Medicine, Baltimore, MD, USA

<sup>32</sup> Department of Orthopaedic Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

<sup>33</sup> Spine Surgery Unit, Department of Aging, Orthopaedic and Rheumatological Sciences, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy

<sup>34</sup> Department of Orthopaedic Surgery, Rothman Institute and Thomas Jefferson University, Philadelphia, PA, USA

<sup>35</sup> Department of Orthopaedics, Ganga Hospital, Coimbatore, India

<sup>36</sup> Department of Neurosurgery, 251 General Air Force and Reserve Hospital, Athens, Greece

<sup>37</sup> Operative Research Unit of Orthopaedic and Trauma Surgery, Fondazione Policlinico Universitario Campus Bio-Medico, Rome, Italy

<sup>38</sup> Research Unit of Orthopaedic and Trauma Surgery, Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Rome, Italy

<sup>39</sup> Department of Orthopaedic Surgery, Noordwest Ziekenhuisgroep, Alkmaar, The Netherlands

<sup>40</sup> Department of Neurological Surgery, Weill Cornell Medicine, New York, NY, USA

<sup>41</sup> Department of Orthopedic Surgery, Columbia University, New York, NY, USA

raise concern that the reportedly excellent fusion rates reported following ACDF surgeries may be artificially inflated due to questionable radiographic methodologies.

Thus, there is need for a standardization in the radiographic assessment of fusion status. Several recent reviews have each concluded that the preferred initial screening method for non-union should be to assess dynamic ISPM, with the cutoff for suspected non-union being more than 1 millimeter (mm) of motion.<sup>2-4</sup> This includes reviews from the research committee of the Cervical Spine Research Society,<sup>3</sup> as well as the Joint Section report from the American Association of Neurological Surgeons and the Congress of Neurological Surgeons.<sup>9</sup>

However, the ability of radiographic reviewers to measure ISPM reliably and reproducibly with a 1 mm accuracy is not clear. The majority of studies have utilized a small number of reviewers (typically 1-3), and correlation coefficients for the agreement between reviewers is infrequently published.<sup>1,3,9</sup> The authors of these studies have typically been senior level members of the major cervical societies, and it is not clear if the skill of the expert reviewers in these papers is reproducible amongst the overall population of spinal surgeons. Furthermore, from a statistical standpoint, correlation coefficients represent an estimate of the variance in the measurement. A high correlation implies the variance amongst reviewers is low, but does not support a specific numeric measurement cutoff. Some studies have used receiver operating characteristics to support a cutoff number for ISPM, but it is not clear if the observed variance in measurements actually supports the level of accuracy necessary for that cutoff to be statistically valid. Thus, the purpose of this study was to determine the minimum detectable distance that could be reliably measured by practicing surgeons during assessment of ISPM on flexion-extension radiographs, and to determine if the previously suggested measurement cutoff of <1 mm ISPM is statistically reproducible amongst practicing surgeons.

## Methods

### Study Design

Institutional Review Board approval was obtained for this study (STUDY00000102). This retrospective review of pre-existing records was deemed to provide no more than minimal risk to subjects and a waiver of patient consent was granted. AO Spine Knowledge Forum Degenerative (KF Degen) consists of 54 surgeons and researchers from over 15 countries around the world. Since 2019, cases from this group have been entered into a prospective clinical registry, which includes radiographic images. We queried this database to identify cases of ACDF performed with an interbody device and an anterior plate performed for a degenerative indication that had 12-month post-operative flexion-extension images available. Cases were excluded if they included circumferential instrumentation, if there were no posterior spinous processes available to measure, or if they were performed for a non-

degenerative indication (trauma, tumor, infection, deformity). Images were also excluded if the spinous processes were not visible due to patient body habitus or due to poor radiographic technique. Lastly, the assessment of motion requires that the patient has put in a good effort in both the extension and flexion images. Prior authors have suggested a cutoff of >4 mm of ISPM at the level adjacent to the fusion to assess adequacy of the motion effort, and thus we excluded images if they did not meet this threshold.<sup>10</sup> Each radiograph was reviewed by a committee of the three senior authors (CM, TY, DR) to verify that it met the inclusion criteria. Images from 9 patients with 19 levels of surgery were included. We elicited participation from amongst KF Degen members, and 29 surgeons agreed to participate. Each participant is a board certified and fellowship trained spinal surgeon who regularly performs anterior cervical fusions.

### Radiological Assessment and Statistical Analysis

Each participant was asked to view a 2-minute training video which discussed the rationale for the study and also demonstrated proper measurement technique ([Supplementary Appendix Video 1](#)). Specifically, each reviewer was instructed to zoom in to maximize the measured level within the field of view on the screen, with a minimum magnification of at least 150%. The contrast on the image was then adjusted to maximize the visible difference between the bony landmarks and the surrounding soft-tissues. Surgeons were next asked to identify an obviously identifiable landmark around the tip of the spinous process, with the landmark being visible on both flexion and extension views simultaneously ([Figure 1](#)).<sup>10</sup> Measurement of ISPM is affected by the distance from the axis of rotation that the measurement is taken. Measurements taken at the very tip of the spinous process are therefore likely to have larger ISPM values than measurements taken near the facet or in the mid aspect of the spinous process. For this reason, reviewers were specifically asked to identify bony landmarks near the tip of the spinous process for measurement and this was also noted in the instructional video. The measurement in millimeters between the spinous process landmarks was recorded on both the flexion and extension image. The ISPM was then determined by subtracting the distance measured on the extension image from the distance measured at the same level, and using the same landmarks, on the flexion image. Each participant was required to attest that they had viewed the video and had also practiced the measurement technique on at least 10 of their own cases prior to participating in the radiographic study.

Each case was de-identified of patient information and assigned a number. Each reviewer was then sent DICOM files of the images and was instructed to load the images onto their local PACS system for measurement. The reviewers also measured each image a second time, at least two weeks after their first measurement, to allow calculation of intra-observer correlation. Inter-observer correlation coefficients (ICC),



**Figure 1.** Flexion (left) and Extension (right) images showing an example of using at least 150% magnification and adjusting the image contrast to identify bony landmarks (marked by arrow) for ISPM measurement.

Intra-observer correlation coefficients, Standard Error of Measurement (SEM), and the Minimum Detectable Difference (MD) were calculated off these measurements. Correlation Coefficient estimates and their 95% confidence intervals (CI) were calculated using a 2-way random-effect repeated-measures analysis of variance (ANOVA). Values of ICC less than .5, between .5 and .75, .75 and .9, and greater than .90 indicate poor, moderate, good, and excellent reliability, respectively.<sup>11</sup> The SEM was derived from the standard deviation (SD) of all motion values and the ICC, ie,  $SEM = SD$ . The MD was then finally derived, based on the SEM using the following formula:  $MD = SEM * 1.96$ .<sup>12</sup> Statistical analyses were performed using the software SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). Numeric variables are presented as mean  $\pm$  standard deviation (SD). Categorical variables are summarized using counts (n) and percentages (%).

Lastly, we asked reviewers to report on the fusion status of the operative level. Reviewers were instructed to use their expert clinical opinion and to rely on a combination of their ISPM measurements as well as other factors such as the absence of bridging and remodeled bone around the interbody, presence of screw or hardware loosening, and radiolucent lines around implants or at the interbody-endplate interface. The reviewer could report that the level was obviously fused, an obvious non-union, or indeterminate. We then calculated ICC for this assessment.

### Clerical and Data Processing Errors

Clerical and data processing errors in research studies are common, occurring in up to 28% of data-entries.<sup>13</sup> Examples of these types of errors in radiographic measurement studies include forgetting to type a decimal thus leading the measurement to be off by an order of magnitude and transcribing a measurement from the wrong level. To identify these errors, surgeons were asked to go back and review their measurements for any level in which their measurement differed from the group median by one absolute deviation or more. If a

clerical error in one of the above categories was identified, then the old measurement was discarded and a new measurement was taken. For transparency, both the initial calculations and then the updated calculations after removal of the clerical errors are both presented here.

### Results

In the first round of measurement the inter-observer ICC was .76, the SEM was 1 mm and the MD was 2.76 mm (Table 1, Figure 2). Intra-observer variability was .71 (.56, .82). Agreement on fusion status was moderate with an inter-observer ICC of .6 (.44; .76). Surgeons were asked to identify gross clerical errors in their first round of measurement. A total of 181 measurements were found to be more than 1 absolute deviation from the group mean and were thus re-evaluated by the original surgeon. 62 of the measurements were changed by  $>1$  mm, 95 further cases were changed 0-1 mm, and for 22 cases the surgeon did not find any error upon reviewing their measurement and thus the number did not change. After revising those errors, the inter-observer ICC improved to .82, the SEM improved to .83 mm and the MD improved to 2.29 mm (Table 1, Figure 3). We performed a sub-analysis of the six highest performing reviewers who had an ICC above .9, and found that the SEM improved to .45 mm and the MD improved to 1.26 mm (Table 1, Figure 4).

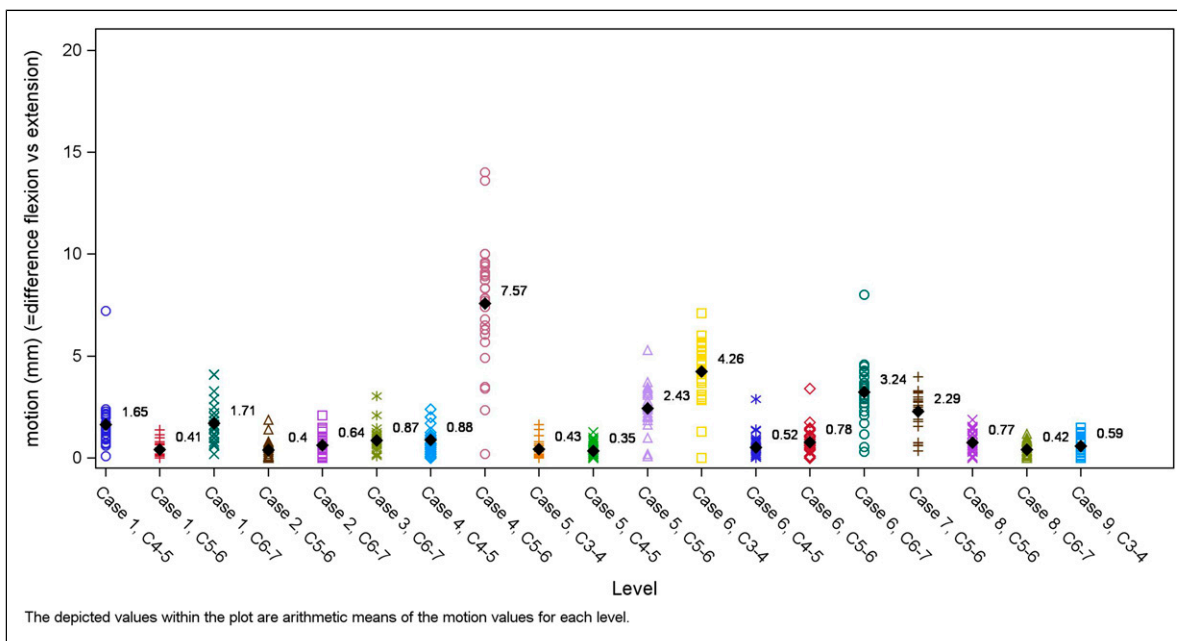
### Discussion

While multiple prior reviews have supported the importance of dynamic imaging in assessing fusion status after ACDF,<sup>2,3,6,7</sup> the specific motion cutoffs have varied substantially across studies. A review from 2018 found 11 prior articles that had published cutoffs for non-union based on ISPM, ranging from 0-3 mm.<sup>4</sup> If fusion is complete both anteriorly in the disc space and also posteriorly through the facets, then motion at that segment should be 0 mm. Based on this assumption, at least six prior studies suggested a cutoff of

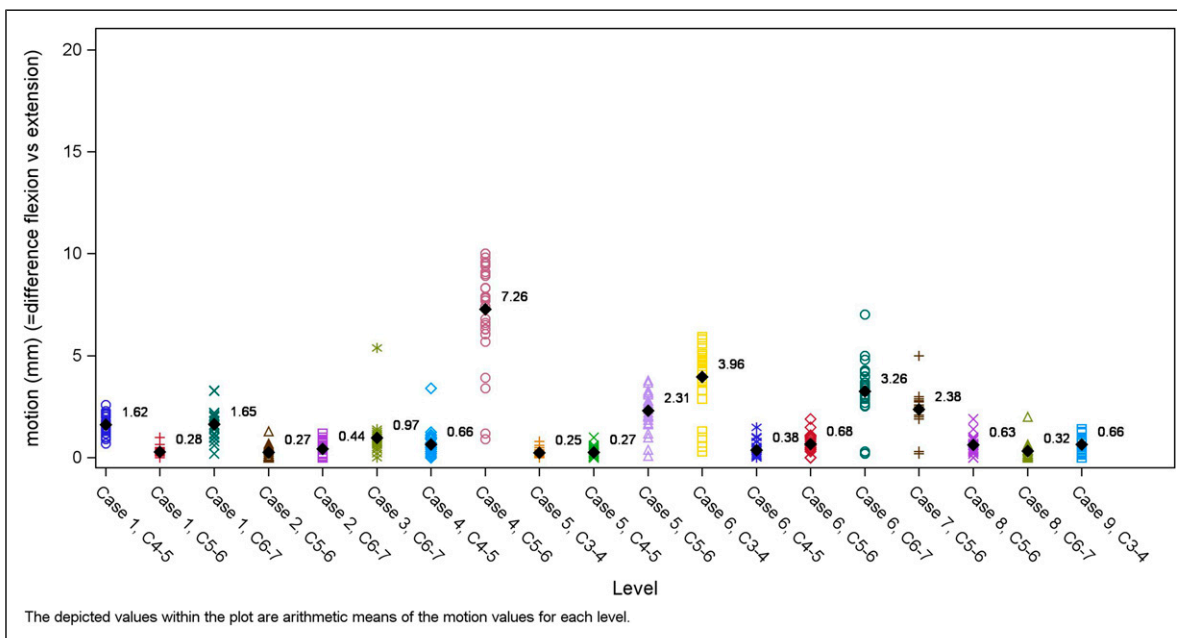
**Table 1.** Statistical Outcomes.

	Interrater CC	SEM (mm)	MD (mm)
All reviewers initial result	.76 (.64; .88)	1.00	2.76
All reviewers after removal of clerical errors	.82 (.71; .91)	.83	2.29
Six expert reviewers	.94 (.89; .97)	.45	1.26

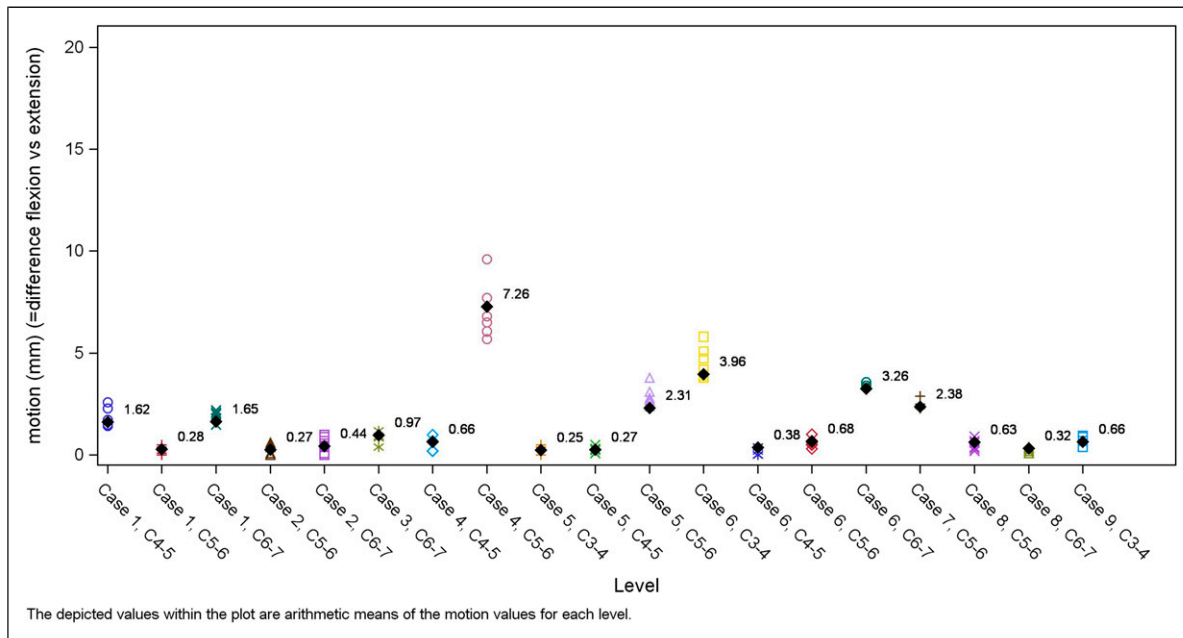
CC, Correlation coefficient; SEM, Standard error of measurement; MD, Minimum detectable difference; mm, millimeter.



**Figure 2.** Measurement of ISPM (mm) from all reviewers in the first round of measurement.



**Figure 3.** Measurement of ISPM (mm) from all reviewers after the removal of clerical errors.



**Figure 4.** Measurement of ISPM (mm) from the six expert reviewers.

0 mm of ISPM for determining non-union.<sup>4</sup> However, requiring strict adherence to this cutoff would mean any motion of .1 mm or more would be considered a non-union, and it seems unlikely that surgeons can reproducibly achieve that measurement precision. Furthermore, some motion may still be observed due to plastic deformation of the fusion mass when the disc space is fused, but the facets remain open, which is common following anterior fusions. Thus, the majority of prior authors have advocated for a cutoff of 1-2 mm, but the validity of these cutoffs is not well established.<sup>1-3,7,10,14</sup>

The gold standard for assessment of fusion status is intraoperative exploration, but few prior studies have compared radiographic techniques against this standard in large numbers.<sup>8,14,15</sup> Song et al reported on 125 patients who had undergone operative exploration with confirmation of fusion status.<sup>10</sup> Radiographs from that series were measured by two independent reviewers, and a receiver operator curve analysis found that the ideal cutoff for assessment of non-union should be an ISPM of .9 mm, with a sensitivity and specificity similar to that from CT scans.<sup>10</sup> The ICC of the two measurers was .825 (95% CI: .782 to .862). Based in part on these results, research committees from both the Cervical Spine Research Society,<sup>3</sup> as well as the Joint Section report from the American Association of Neurological Surgeons and the Congress of Neurological Surgeons,<sup>9</sup> have concluded that assessment of ISPM on dynamic radiographs should be the initial screening study for non-union, with a suggested cutoff of 1 mm.

However, the studies cited by the research committee reviews typically report only CC values for the observer's measurements, and do not report on the minimum detectable difference (MD). Receiver operating characteristics of ISPM measurements have been used to support a specific cutoff

of <1 mm,<sup>10</sup> but the ability of surgeons to accurately measure ISPM <1 mm is not clear. Thus, in this study, we aimed to determine the MD of ISPM measurements amongst practicing spine surgeons. Calculation of the MD represents the measurement distance at which it can be statistically certain that two measurements are different from each other, and takes into account both the precision of the number and also the reproducibility of the number across the multiple reviewers. Measurements less than the MD may be different simply due to measurement technique, and may not represent a true difference.

The ICC's reported here (.76-.82) are high and similar to those reported by other studies,<sup>10</sup> which indicates that the ISPM for each level can be reliably distinguished. However, the MD was 2.29 mm, which indicates that the precision of the measured values was not high enough to support a statistical cutoff of <1 mm. In other words, an average radiographic reviewer may not have enough precision to reliably and consistently identify measurement differences <1 mm. Furthermore, the ICC agreement on fusion status was only .6, which implies moderate variability. This indicates that the surgeons in this study did not have consistently reliable assessments of fusion status based on dynamic radiographs alone, and implies that prior studies which have relied solely on plain radiographs for fusion assessment may not be accurately classifying the non-unions.

There could be several reasons for this. First, surgeons participating in a research study in which they have to take a large number of measurements may become fatigued over the course of taking many measurements, leading to a decrease in their precision. In contrast, in a clinical setting, the surgeon only needs to measure one patient's radiographs at a time, and

that measurement is likely to have immediate clinical relevance. Therefore, surgeons are likely to take more time or take greater care, and that may result in better precision than we observed in this research setting. Second, it is clear that the quality of measurements varied across reviewers. We performed a sub-analysis of the 6 highest performing reviewers who had ICC >.90 and found a MD of 1.26 mm, which is significantly more precise than the overall group MD of 2.29 mm. We asked these reviewers about their technique, and they report taking significant time and care in the measurements, in some cases up to 10 minutes per measurement, as opposed to 1 minute or less for most reviewers. This implies that the technique is quite reliable in the hands of experts who take great care in making the measurement. Thus, although the group average ISPM MD of 2.29 was above the common clinical threshold of 1 mm, we believe our results do support the use of ISPM as a radiographic technique for expert reviewers. Bringing the image into maximum magnification (at least 150% or greater) and using the contrast tool to identify clear and reproducible bony landmarks were common strategies used by the best performing reviewers to increase the precision of their technique (Figure 1).

Lastly, in the initial round of statistical analysis we identified several instances of clerical errors, including examples of transcribing a measurement from the wrong level, forgetting a decimal, or typing a wrong number by mistake. We believe that this supports our hypothesis that fatigue and imprecision played a significant role in our results; few surgeons would make the mistake of calling an interspinous process motion of 2 mm as being 2 centimeters (cm) in a clinical setting, since 2 cm would be grossly abnormal. A prior systematic review reported that data-processing errors of this kind are common, occurring in up to 28% of data entries (up to 2784 errors per 10 000 data-fields).<sup>13</sup> Each reviewer in this study sent us their measurements in a spread-sheet, which meant that we could not directly view how they took the measurement. We identified some of these errors by asking each reviewer to go back and check their measurement whenever it was more than one absolute deviation greater than the group mean. However, this method is imprecise, and may have missed some clerical errors that were less than one deviation from the mean. Given the prevalence of these data processing errors both in our study, as well as the reportedly high rate from prior review,<sup>13</sup> we suggest that future studies have some method of auditing the measurement itself. Specifically, a screenshot of the measurement could be saved and sent along with the spreadsheet so that if questions arise the image can be directly audited.

In contrast to prior studies that have utilized 1-3 radiographic reviewers, the primary strength of this study is the recruitment of 29 practicing spine surgeons from around the world, and its use of statistical techniques to calculate the minimum detectable difference in surgeon measurements. The reviewers utilized the imaging software at their home

institution to make the measurements, which is meant to reproduce a real-world clinical scenario for how the measurements are regularly taken.

In summary, radiographic assessment of ISPM by human reviewers can be accurate in the hands of expert reviewers who take great care in the measurements, and approaches the precision needed to be statistically certain when using a motion cutoff of <1 mm. However, the average reviewer in this study was less precise with an MD of 2.29 mm, and agreement on fusion status was only moderate. Thus, an average reviewer may not be able to rely on dynamic radiographs alone for assessment of fusion status, and may benefit from additional imaging such as CT scans, or AI based computer aids. Prior studies that rely solely on dynamic radiographs may have inaccurately classified some cases of non-unions. Clerical errors can occur in data transcription, and future studies would benefit from incentivizing precise reading of radiographs as well as employing a method that allows auditing of the measurements for accuracy.

## Acknowledgments

This study was organized and supported by AO Spine through the AO Spine Knowledge Forum Degenerative, a focused group of international spine experts.

## 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.

## Ethical Statement





### Ethical Approval

Institutional Review Board approval was obtained for this study. Emory University (AM1\_IRB00102479), University of Minnesota (STUDY00000102), Mount Sinai Health System (18-0052), Ärztekammer Sachsen-Anhalt – K örüerschaft des öffentlichen Rechts (59/16).

## ORCID iDs

Sangwook Tim Yoon  <https://orcid.org/0000-0003-1010-6952>  
Ram Kiran Alluri  <https://orcid.org/0000-0001-5919-707X>  
Samuel K. Cho  <https://orcid.org/0000-0001-7511-2486>  
Xiaolong Chen  <https://orcid.org/0000-0002-8128-0969>  
Juan P. Cabrera  <https://orcid.org/0000-0003-4685-6106>  
Stipe corluka  <https://orcid.org/0000-0002-5732-1150>  
Andreas K. Demetriades  <https://orcid.org/0000-0002-2004-9448>  
Zohar Ghogawala  <https://orcid.org/0000-0003-1345-1831>  
Wael Hamouda  <https://orcid.org/0000-0002-4627-0992>  
Inbo Han  <https://orcid.org/0000-0002-0834-9325>  
Patrick C. Hsieh  <https://orcid.org/0000-0002-7206-4842>



Amit Jain  <https://orcid.org/0000-0002-9983-3365>  
 Hai V. Le  <https://orcid.org/0000-0002-9111-9060>  
 Philip K. Louie  <https://orcid.org/0000-0002-4787-1538>  
 Hans-Jorg Meisel  <https://orcid.org/0000-0003-3838-1489>  
 Sathish Muthu  <https://orcid.org/0000-0002-7143-4354>  
 Laura Scaramuzzo  <https://orcid.org/0000-0001-6570-121X>  
 Shanmuganathan Rajasekaran  <https://orcid.org/0000-0001-6043-006X>  
 Yabin Wu  <https://orcid.org/0000-0002-1836-7674>  
 K. Daniel Riew  <https://orcid.org/0000-0002-0718-3423>

## References

- Lin W, Ha A, Boddapati V, Yuan W, Riew KD. Diagnosing pseudoarthrosis after anterior cervical discectomy and fusion. *Neurospine*. 2018;15(3):194-205. doi:10.14245/ns.1836192.096
- Riew KD, Yang JJ, Chang DG, et al. What is the most accurate radiographic criterion to determine anterior cervical fusion? *Spine J*. 2019;19(3):469-475. doi:10.1016/j.spinee.2018.07.003
- Rhee JM, Chapman JR, Norvell DC, Smith J, Sherry NA, Riew KD. Radiological determination of postoperative cervical fusion: a systematic review. *Spine*. 2015;40(13):974-991. doi:10.1097/BRS.0000000000000940
- Oshina M, Oshima Y, Tanaka S, Riew KD. Radiological fusion criteria of postoperative anterior cervical discectomy and fusion: a systematic review. *Global Spine J*. 2018;8(7):739-750. doi:10.1177/2192568218755141
- Wang Z, Zhu R, Yang H, et al. Zero-profile implant (Zero-p) versus plate cage benezech implant (PCB) in the treatment of single-level cervical spondylotic myelopathy. *BMC Musculoskelet Disord*. 2015;16:290. doi:10.1186/s12891-015-0746-4. Published 2015 Oct 12.
- Park DK, Rhee JM, Kim SS, Enyo Y, Yoshiok K. Do CT scans overestimate the fusion rate after anterior cervical discectomy and fusion? *J Spinal Disord Tech*. 2015;28(2):41-46. doi:10.1097/BSD.0b013e31829a37ac
- Lambrechts MJ, D'Antonio ND, Karamian BA, et al. What is the role of dynamic cervical spine radiographs in predicting pseudoarthrosis revision following anterior cervical discectomy and fusion? *Spine J*. 2022;22(10):1610-1621. doi:10.1016/j.spinee.2022.04.020
- Ghiselli G, Wharton N, Hipp JA, Wong DA, Jatana S. Prospective analysis of imaging prediction of pseudoarthrosis after anterior cervical discectomy and fusion: computed tomography versus flexion-extension motion analysis with intraoperative correlation. *Spine*. 2011;36(6):463-468. doi:10.1097/BRS.0b013e3181d7a81a
- Kaiser MG, Mummaneni PV, Matz PG, et al. Radiographic assessment of cervical subaxial fusion. *J Neurosurg Spine*. 2009;11(2):221-227. doi:10.3171/2009.3.SPINE08719
- Song KS, Piyaskulkaew C, Chuntarapas T, et al. Dynamic radiographic criteria for detecting pseudoarthrosis following anterior cervical arthrodesis. *J Bone Joint Surg Am*. 2014;96(7):557-563. doi:10.2106/JBJS.M.00167
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research [published correction appears in *J chiropr med*. 2017 dec;16(4):346]. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
- Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240. doi:10.1519/15184.1
- Garza MY, Williams T, Ounpraseuth S, et al. Error rates of data processing methods in clinical research: a systematic review and meta-analysis of manuscripts identified through PubMed. Preprint. *Res Sq*. 2023:rs.3.rs-2386986. doi:10.21203/rs.3.rs-2386986/v2. Published 2023 Dec 21.
- Cannada LK, Scherping SC, Yoo JU, Jones PK, Emery SE. Pseudoarthrosis of the cervical spine: a comparison of radiographic diagnostic measures. *Spine*. 2003;28(1):46-51. doi:10.1097/00007632-200301010-00012
- Buchowski JM, Liu G, Bunmaprasert T, Rose PS, Riew KD. Anterior cervical fusion assessment: surgical exploration versus radiographic evaluation. *Spine*. 2008;33(11):1185-1191. doi:10.1097/BRS.0b013e318171927c