

## Heterotopic Ossification After Cervical Total Disc Replacement at 7 Years—Prevalence, Progression, Clinical Implications, and Risk Factors

PIERCE D. NUNLEY, DAVID A. CAVANAUGH, EUBULUS J. KERR III, PHILLIP ANDREW UTTER, PETER G. CAMPBELL, KELLY A. FRANK, KYLE E. MARSHALL and MARCUS B. STONE

*Int J Spine Surg* 2018, 12 (3) 352-361

doi: <https://doi.org/10.14444/5041>

<https://www.ijssurgery.com/content/12/3/352>

This information is current as of May 17, 2025.

---

**Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at:  
<http://ijssurgery.com/alerts>

# Heterotopic Ossification After Cervical Total Disc Replacement at 7 Years—Prevalence, Progression, Clinical Implications, and Risk Factors

PIERCE D. NUNLEY, MD, DAVID A. CAVANAUGH, MD, EUBULUS J. KERR III, MD, PHILLIP ANDREW UTTER, MD, PETER G. CAMPBELL, MD, KELLY A. FRANK, MS, KYLE E. MARSHALL, MS, MARCUS B. STONE, PhD

*Spine Institute of Louisiana, Shreveport, Louisiana*

## ABSTRACT

**Background:** Heterotopic ossification (HO) is a known risk following cervical total disc replacement (CTDR) surgery, but the cause and effect of HO are not well understood. Reported HO rates vary, and few studies are specifically designed to report HO. The effects on outcomes, and the risk factors for the development of HO have been hypothesized and reported in small-population, retrospective analyses, using univariate statistics.

**Methods:** Posthoc, multiple-phase analysis of radiographic, clinical, and demographic data for CTDR as it relates to HO was performed. HO was radiographically graded for 164 one-level and 225 two-level CTDR patients using the McAfee and Mehren system. Analysis was performed to correlate HO grades to clinical outcomes and to evaluate potential risk factors for the development of HO using demographics and baseline clinical measures.

**Results:** At 7 years, 1-level clinically relevant HO grades were 17.6% grade 3 and 11.1% grade 4. Two-level clinically relevant HO grades, evaluated using the highest patient grade, were 26.6% grade 3 and 10.8% grade 4. Interaction between HO and time revealed significance for neck disability index (NDI;  $P = .04$ ) and Visual Analog Scale (VAS) neck pain ( $P = .02$ ). When analyzed at each time point NDI was significant at 48–84 months and VAS neck at 60 months. For predictors 2 analyses were run; odds ratios indicated follow-up visit, male sex, and preoperative VAS neck pain are related to HO development, whereas hazard ratios indicated male sex, obesity, endplate coverage, levels treated, and preoperative VAS neck pain.

**Conclusions:** This is the largest study to report HO rates, and related outcomes and risk factors. To develop an accurate predictive model, further large-scale analyses need to be performed. Based on the results reported here, clinically relevant HO should be more accurately described as motion-restricting HO until a definitive link to outcomes has been established.

TDR

Keywords: cervical total disc replacement, heterotopic ossification, Mobi-C

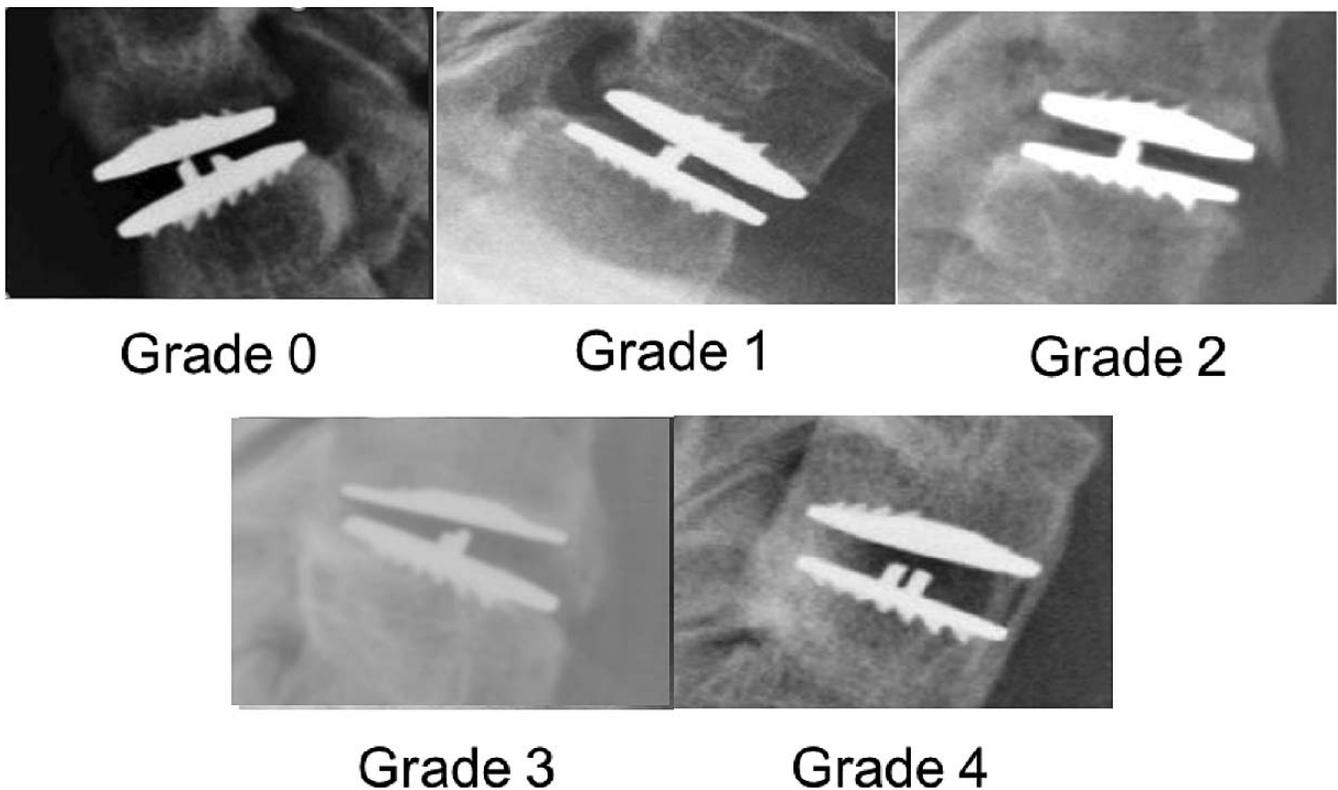
## INTRODUCTION

The introduction of cervical total disc replacement (CTDR) began with the concept of motion preservation to treat degenerative disc disease (DDD). Motion preservation most closely mimics natural motion of the spine and is believed to preserve the adjacent segments from degeneration over the long term compared with cervical anterior discectomy and fusion (ACDF).<sup>1–4</sup> CTDR has been well studied, with multiple US Food and Drug Administration approvals, and long-term clinical trial data indicate it is a safe and effective treatment for cervical DDD.<sup>5–15</sup>

An unintended and not well understood sequelae of CTDR is heterotopic ossification (HO). Follow-

ing CTDR, HO can develop around the device between vertebral bodies. In severe cases, HO limits range of motion, sometimes creating fusion of the segment.<sup>16–19</sup> Reported rates of HO following CTDR vary drastically, creating more debate and concern around the true rate and impact of HO.<sup>20–23</sup> The long-term effects of HO resulting in unintended fusion have not been analyzed.

Analysis of HO risk factors began in hip arthroplasty with studies indicating sex differences in HO development.<sup>24,25</sup> For CTDR, HO risk factors have been hypothesized, including lack of nonsteroidal anti-inflammatory drug (NSAID) use postoperatively, sex, age, surgical level, number of treated levels, preoperative degeneration, and sur-



**Figure 1.** Representative grades of heterotopic ossification.

gical technique. Although there are multiple studies that attempt to correlate CTDR HO to risk factors, these studies are limited in scope due to small patient populations (largest is 170 patients), retrospective data, and univariate statistical analyses.<sup>18,26–31</sup> Results of these studies are varied and sometimes contradictory, so the question of risk factors for development of HO remains unanswered. This study includes the largest known prospectively collected CTDR data set analyzed with multiple statistical models as a first step to identifying risk factors associated with HO.

## MATERIALS AND METHODS

Data were collected as part of a prospective, randomized, controlled, multicenter clinical trial comparing CTDR (Mobi-C, Zimmer Biomet, Austin, Texas) with ACDF for treatment of symptomatic DDD with radiculopathy or myelodysplasia at 1 or 2 contiguous levels from C3 to C7. ACDF data collected were not used for this analysis. Enrollment for the CTDR study arm included 164 one-level and 225 two-level patients. Details of the clinical trial have been reported previously.<sup>13,15</sup>

## Study Design

This study was designed as a posthoc, multiple-phase analysis of radiographic, clinical, and demographic data for 1- and 2-level CTDR as it relates to HO.

## Radiographic Analysis

Available radiographs were analyzed annually from years 1 to 5 and at 7 years postoperatively. Radiographic evaluations were performed by independent radiologists (MMI Inc., Houston, Texas).

HO was graded using the adapted McAfee and Mehren system by 2 independent US-based, board-certified, fellowship-trained, practicing radiologists.<sup>32</sup> In the event of disagreement, a third reader adjudicated their assessments. Representative HO grades are shown in Figure 1, and the protocol definitions of HO are included as Table 1. Grades 0, 1, and 2 were classified as not clinically relevant, whereas grades 3 and 4 HO were classified as clinically relevant.

Range of motion was measured and reported preoperatively and at last follow-up.

**Table 1.** Grading scale of HO.

HO	HO Will Be Graded as “Grade 0, I, II, III, or IV” in Accordance With the Following Definitions
0: grade 0	No evidence of osteophyte formation or heterotopic ossification.
1: grade 1	HO is detectable in the front or sides or the vertebral body, or as islands of bone in the adjacent soft tissue, but it is not in the intervertebral disc space. Bone is not present between the planes formed by the 2 vertebral endplates.
2: grade 2	HO is growing into the disc space. Bone is present between the planes formed by the 2 adjacent endplates but is not significantly blocking or articulating between adjacent vertebral endplates or osteophytes.
3: grade 3	The ROM of the vertebral endplates is blocked by the formation of HO and/or postoperative osteophytes on flexion-extension radiographs, but some movement of the prosthesis still remains.
4: grade 4 (bridging)	HO is causing bony ankylosis. An apparent continuous connection of bridging bone exists between the adjacent vertebral endplates, with little or no motion occurring across the treated segment.

Abbreviations: HO, heterotopic ossification; ROM, range of motion.

### Postoperative Care

The protocol under which the data were initially collected included specific provisions for postoperative NSAID use. The protocol prohibited the use of NSAIDs for both CTDR and ACDF from 1 week preoperatively to 3 months postoperatively, unless specifically prescribed to treat HO.

### Baseline Characteristics and Clinical Outcomes

Demographics, baseline measures, and clinical outcomes were collected prospectively in the trial. Posthoc analysis was performed to correlate grades of HO to patient clinical outcomes and to evaluate potential risk factors for development of grade 3 or 4 HO using demographics and baseline clinical measures. Demographics and baseline clinical scores included age, sex, body mass index (BMI), segmental range of motion (ROM) in flexion/extension, visual analog scale (VAS) neck pain, and disc space height.

Patient clinical outcomes included neck disability index (NDI), VAS neck/arm pain, and Short Form 12 (SF-12) Physical Component Score (PCS)/Mental Component Score (MCS).

### Statistical Methods

Mixed-effects analysis of variance (ANOVA) was used to assess the impact of clinically relevant HO on NDI, VAS neck/arm pain, and SF-12 PCS/MCS scores. A model was created for each of the 5 patient-reported outcomes. For 2-level patients, HO

status was classified by the highest grade of HO at either level. The effect of clinically relevant HO was adjusted for follow-up visit, number of levels treated, BMI, age, and sex, with random effects for study site and patients within study sites. An interaction term for HO status and follow-up visit was included to determine whether the effect of HO differed across time. If the interaction term was significant, the HO effect was tested at each follow-up visit. Type 3 tests of fixed effects were used to calculate *P* values.

Significant predictors of clinically relevant HO were identified by mixed-effects logistic regression. Features were chosen based on exploratory analysis and those believed to have potential relevance to HO. Preoperative features included age, sex, BMI, ROM in flexion/extension, VAS neck pain, and disc space height. Postoperative predictors included follow-up time, endplate coverage, and number of levels treated. Endplate coverage was calculated as the average percent of anterior-posterior device contact of both inferior and superior endplates. For 2-level patients, each level was treated as a separate observation. Random effects for individual patients and level within patients were included to adjust for within-patient and within-level dependencies across multiple follow-up observations. Missing data were imputed using multivariate chained equations. *P* values and 95% confidence intervals were calculated using Wald approximation.

Additionally, a mixed-effects Cox proportional hazards model was fit with the same features as the previous logistic model. This was done to (1) validate the robustness of the logistic model results and (2) express variable effects as hazard ratios, which are more readily interpretable as relative probability ratios for obtaining clinically relevant HO at any given point.

A *P* value threshold of 0.05 was used to determine statistical significance of model parameter estimates. Statistical computations were performed in R version 3.2.4 (R Foundation for Statistical Computing, Vienna, Austria).

## RESULTS

### Radiographic Analysis

HO data were available for 65.9% (108 of 164) of 1-level and 70.2% (158 of 225) of 2-level patients at 7 years. Grades of HO at each time point are

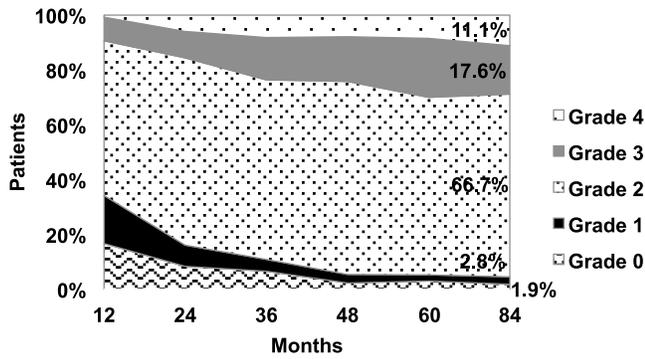


Figure 2. One-level heterotrophic ossification by grade.

included as Figures 2, 3, and 4. Two-level HO is presented separately for superior and inferior levels.

One-level clinically relevant HO grades were 28.7%, with 17.6% (grade 3) and 11.1% (grade 4). The nonrelevant HO included 1.9% (grade 0), 2.8% (grade 1), and 66.7% (grade 2). Two-level clinically relevant HO grades were 23.3% superior and 31.7% inferior, with inferior values of 27.0% (grade 3) and 4.7% (grade 4), and superior values of 16.8% (grade 3) and 6.5% (grade 4). The nonrelevant HO for the superior segment included 1.9% (grade 0), 2.6% (grade 1), and 72.3% (grade 2), whereas the inferior segment included 1.4% (grade 0), 2.0% (grade 1), and 64.9% (grade 2). The grades of HO progressed over time, but they slowed after 5 years. From years 5 to 7, only 6.6% of 1-level and 15.2% of 2-level patients experienced a progression of 1 grade or more.

Figure 5 illustrates 2-level HO per patient using the highest grade of HO present. Occurrence of grade 4 HO at 7 years between 1-level (11.1%) and 2-level (10.8%) patients was similar.

ROM at the last measured HO was negatively correlated with HO grade (Figures 6 and 7), confirming grades 3 and 4 HO were associated with

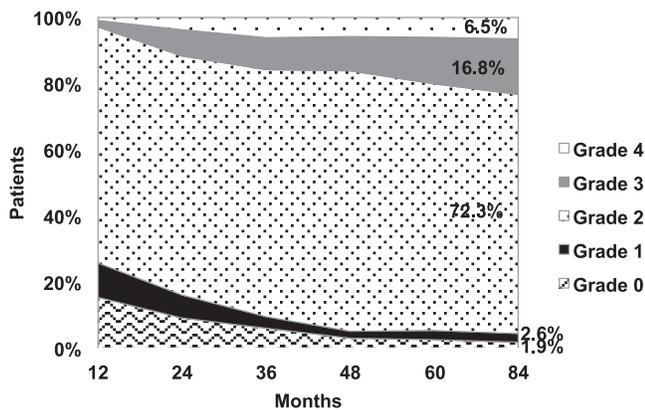


Figure 3. Two-level heterotrophic ossification by grade—superior level.

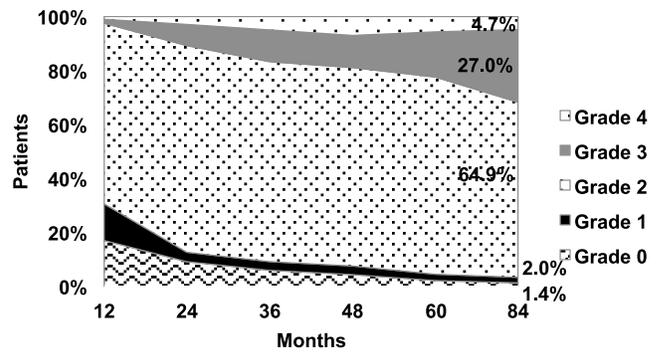


Figure 4. Two-level heterotrophic ossification by grade—inferior level.

restricted segment ROM (1-level: Spearman  $\rho = -0.63$ ,  $P < .0001$ ; and 2-level: Spearman  $\rho = -0.53$ ,  $P < .0001$ ).

### Baseline Characteristics and Clinical Outcomes

Patient baseline characteristics and scores were sorted into clinically relevant and nonrelevant HO present at 7 years (Supplemental Table 1). These results were categorized to evaluate for trends; therefore, statistics were not run on these data.

### Progression of HO and Clinical Outcomes

Results comparing patients with and without clinically relevant HO are presented in Table 2. Clinically relevant HO had no significant effect for all outcomes, when generalized across time. Interaction between HO and follow-up visit was significant for NDI ( $P = .04$ ) and VAS neck pain ( $P = .02$ ), indicating the effect of HO on NDI and VAS neck pain is time dependent and should be analyzed separately at each follow-up period.

For NDI scores, no significant differences between HO groups were observed at early follow-up (12–36 months). At later times (48–84 months), group NDI scores displayed evidence of divergence, with significantly higher scores in the clinically

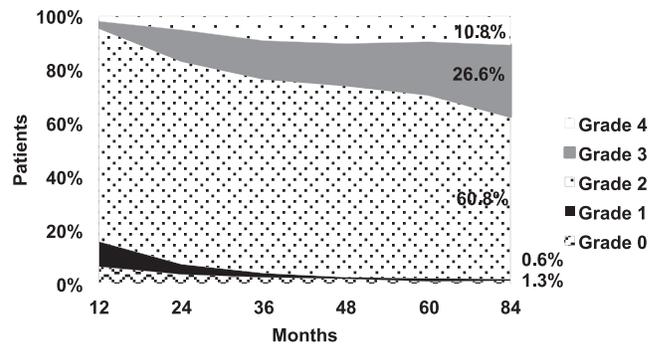


Figure 5. Two-level heterotrophic ossification by grade—by patient.

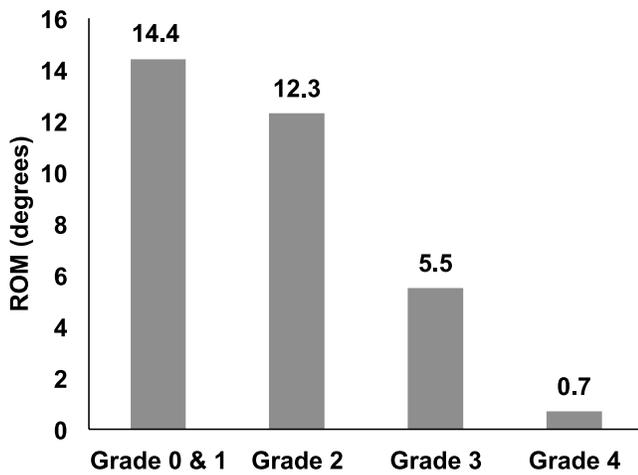


Figure 6. One-level range of motion (ROM) at last follow-up by heterotopic ossification grade.

relevant group (Figure 8 and Table 3). VAS neck pain scores trended higher for the clinically relevant group at later time points, but only month 60 was statistically significant (Figure 9 and Table 3). Sex was a significant effect for NDI, VAS neck/arm pain, and SF-12 PCS scores, with women trending higher pain and function scores.

Clinical Predictors of Clinically Relevant HO

Mixed-effects logistic regression was used to identify significant predictor variables for clinically relevant HO development. Table 4 includes the resulting odds ratio (OR) estimates and associated 95% confidence intervals (CIs). Longer follow-up,

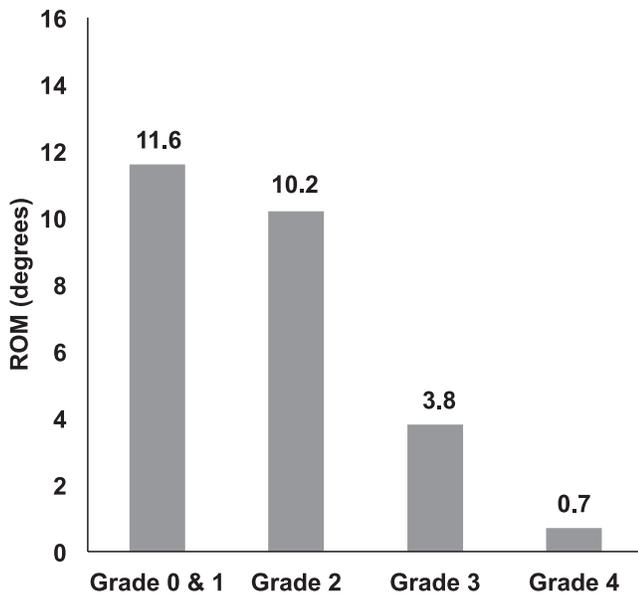


Figure 7. Two-level range of motion (ROM) at last follow-up by heterotopic ossification grade.

Table 2. Results from mixed-model ANOVA.

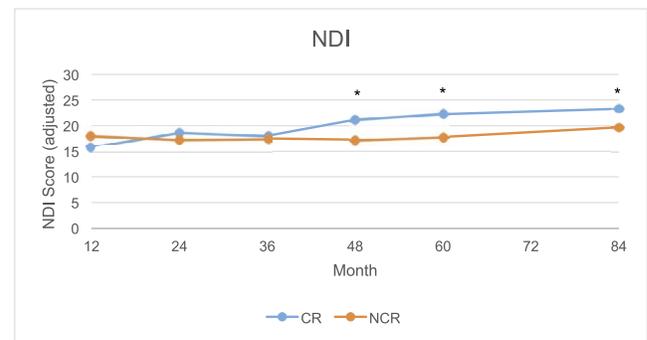
Outcome	Adjusted Mean for NCR HO	Adjusted Mean for CR HO	Adjusted Mean Difference (CR HO – NCR HO)	P Value
NDI	17.9	19.9	2.0	.13*
VAS neck pain	19.2	21.1	1.9	.28*
VAS arm pain	15.1	16.5	1.4	.41
SF-12 PCS	46.7	45.8	-0.9	.30
SF-12 MCS	50.9	50.1	-0.8	.32

Abbreviations: NCR, not clinically relevant; CR, clinically relevant; HO, heterotopic ossification; NDI, neck disability index; VAS, visual analog scale; SF-12, Short Form 12; PCS, Physical Component Score; MCS, Mental Component Score.

\*Indicates a significant interaction between the effect of HO and follow-up visit. Because the interaction term is significant, the P value cannot be readily interpreted. See Table 3 for the effect of HO at each individual time point.

male sex, and higher preoperative VAS neck pain scores demonstrated significantly higher odds of clinically relevant HO. Obesity was associated with higher odds for clinically relevant HO, whereas increasing endplate coverage and 2-level treatment were associated with lower odds of HO; however, these correlations did not reach statistical significance. Preoperative age, ROM, and disc height did not demonstrate a definitive relationship with HO development.

The features from the logistic model were used to fit the mixed-effects Cox proportional hazard model. Cox regression estimates model parameters in terms of hazard ratios, as opposed to logistic regression ORs. Results from the Cox regression (Table 5) largely agreed with the logistic regression analysis. When adjusting for other variables, men were approximately 3 times more likely to develop clinically relevant HO than women ( $P < .0001$ ). Probability of developing HO was inversely related to endplate coverage; that is, a 5% decrease in endplate coverage resulted in 1.3 times higher risk for developing clinically relevant HO ( $P = .004$ ).



\*p < 0.05

Figure 8. Adjusted mean neck disability index (NDI) scores at annual follow-up visits for patients with clinically relevant and non-clinically relevant heterotopic ossification. Error bars present SEM. \*P < .05. Abbreviations: CR, clinically relevant; NCR, not clinically relevant.

**Table 3.** Adjusted NDI and VAS neck pain scores by follow-up visit.

Month	NDI Score				VAS Neck Pain Score			
	NCR	CR	Mean Difference (CR – NCR)	P Value	NCR	CR	Mean Difference (CR – NCR)	P Value
12	18.0	15.9	-2.1	.40	19.9	13.0	-6.9	.11
24	17.3	18.6	1.3	.47	17.7	19.1	1.4	.64
36	17.5	18.0	0.5	.78	19.1	18.6	-0.44	.87
48	17.2	21.2	4.0	.0177	18.4	22.9	4.5	.09
60	17.8	22.2	4.4	.0054	18.6	26.5	7.9	.0019
84	19.7	23.3	3.6	.0312	21.3	26.2	4.9	.07

Abbreviations: NDI, next disability index; VAS, visual analog scale; NCR, not clinically relevant; CR, clinically relevant.

Number of levels treated was inversely related to HO development with a 2-level patient half as likely to develop clinically relevant HO as a 1-level counterpart ( $P = .012$ ). Preoperative VAS neck pain was also related to HO development, with a 30-point increase on the VAS neck pain scale resulting in 1.6 times higher risk ( $P = .008$ ). Obesity was a significant predictor, with obese patients being 3 times more likely to develop HO ( $P = .048$ ). Preoperative age, ROM, and disc height were not significant predictors of clinically relevant HO.

### DISCUSSION

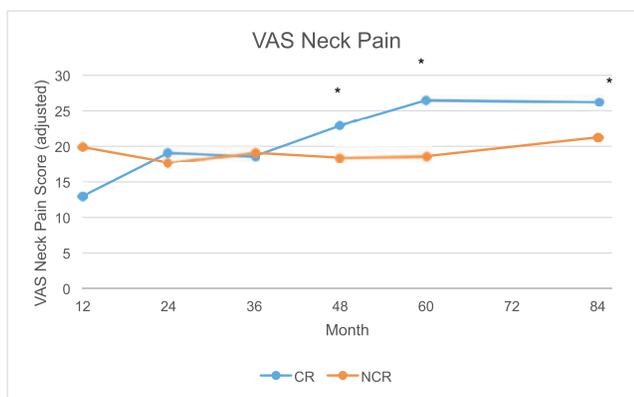
Understanding potential risk factors for developing HO is critical to assisting surgeons with proper patient selection. The ability to accurately predict long-term patient outcomes would be an invaluable tool. For this analysis, we used multiple modeling techniques to validate the consistency of the results and to determine the most accurate model for future analyses.

This study constitutes the largest published single-source results of HO following CTDR surgery, and it includes 1- and 2-level patient

cohorts with radiographic HO data. Radiographic HO data, sometimes viewed as a limitation, was valuable for comparison across multiple studies and devices. Also, in 2011 Tu et al<sup>30</sup> published a study that graded HO using CT and plain radiograph in 36 patients. They reported that radiographs had 80% sensitivity and 88.89% specificity in HO detection. Agreement between the CT grading and radiographic grading had an intraclass correlation coefficient of 0.822 (95% CI, 0.710–0.894,  $P < .001$ ). Grades 3 and 4 were the same when graded by CT and radiograph.

Radiographic reported grade 4 HO per patient (1-level, 11.1%; and 2-level, 10.8%) was similar to HO reported in the literature, although long-term follow-up is not available in all other studies. Multiple meta-analyses and individual device reports have been published on CTDR HO, with minimum 2-year follow-up. Grades of HO using the original McAfee scale include a broad range: grade 0, 12.0%–92.3%; grade 1, 0%–37.4%; grade 2, 2.6%–35.5%; grade 3, 0%–45.0%; grade 4, 0.4%–22.2%.<sup>20–22</sup> A review by Kang et al<sup>23</sup> included a pooled incidence across multiple studies of 27.7% grade 3 and 7.8% grade 4 HO.

Yi et al<sup>19</sup> published a study on HO across multiple CTDR devices. The rates of HO are high, but the study does not differentiate the severity of



\* $p < 0.05$

**Figure 9.** Adjusted mean visual analog scale (VAS) neck pain scores at annual follow-up visits for patients with clinically relevant (CR) and not clinically relevant (NCR) heterotrophic ossification. Error bars present SEM. \* $P < .05$ .

**Table 4.** ORs and 95% CIs of fixed effects from the logistic regression model.

Variables	OR	95% CI, LB, UB	P Value
Age (y)	1.02	0.95, 1.1	.57
Follow-up visit (y)	2.59	2.3, 2.9	<.0001
Sex (male)	12.66	3.30, 48.6	.0002
Obese (BMI $\geq 30$ )	2.68	0.72, 10.0	.14
Endplate coverage (%)	0.96	0.89, 1.04	.31
Levels treated (two)	0.28	0.08, 1.03	.06
Preoperative ROM (degree)	0.95	0.85, 1.06	.37
Preoperative VAS neck pain (score)	1.03	1.0002, 1.06	.048
Preoperative disc height (mm)	1.17	0.61, 2.23	.64

Abbreviations: LB, lower bound; UB, upper bound; BMI, body mass index; ROM, range of motion; VAS, visual analog scale.

**Table 5.** Hazard ratios and 95% CIs of fixed effects from the Cox model.

Variables	Hazard Ratio	95% CI, LB, UB	P Value
Age (year)	1.02	0.99, 1.05	.27
Gender (male)	2.95	1.73, 5.06	<.0001
Obese (BMI $\geq$ 30)	1.68	1.004, 2.81	.048
Endplate coverage (%)	0.95	0.91, 0.98	.004
Levels treated (two)	0.51	0.30, 0.86	.012
Preoperative ROM (degree)	0.95	0.90, 1.01	.08
Preoperative VAS neck pain (score)	1.02	1.004, 1.03	.008
Preoperative disc height (mm)	1.15	0.86, 1.53	.35

Abbreviations: LB, lower bound; UB, upper bound; BMI, body mass index; ROM, range of motion; VAS, visual analog scale.

HO nor does it define clinical impacts.<sup>19</sup> There are other studies that either do not use the McAfee scale, or do not report individual HO grades; these studies are not discussed in our report.<sup>33–35</sup>

The limited long-term data available include 10-year follow-up of the Bryan cervical disc in China. The study included 42 patients with 10-year follow-up; HO rates using the McAfee and Mehren scale were 2.4% grade 2, 33.3% grade 3, and 33.3% grade 4 HO.<sup>36</sup>

Using a modified McAfee scale, Suchomel et al<sup>37</sup> reported on 54 patients treated with ProDisc-C.<sup>22,37</sup> At 4-year follow-up, 50 patients, with 60 implants, had HO rates of 37% for grades 0–2, 45% for grade 3, and 18% for grade 4. The authors believed these HO rates were higher than anticipated, and they offer no statistical analysis of potential predictors or impact on clinical outcomes. However, the limited sample size would restrict a robust analysis of this nature.

Although not all US investigational device exemption studies report HO in the literature, the rates that are reported are on the lower end when compared to other studies.

Although rates of HO vary, HO is a result of CTDR surgery, so we sought to gain a better understanding of potential predictors of HO development and its impact on outcomes. Published studies that analyzed outcomes have shown that clinically relevant HO does not impact clinical outcomes, but the statistical methods have been less sophisticated than this study.<sup>20,38</sup> The use of a mixed-effects ANOVA to evaluate outcomes was important because it allowed the effect of clinically relevant HO to be adjusted for imbalances in covariates, such as BMI and sex, and allowed us to model HO effects over time. Time is a natural confounder for HO on pain and function scores, because as time progresses, HO rates increase, and with time we witness a rise in patient outcome scores. Although significance was found for NDI (2

points) and VAS neck pain (1.9 points) at later time points, the differences did not reach clinical significance.<sup>39</sup> Effects of clinically relevant HO on patient outcomes remains unclear, although this analysis is the first to use modeling methodology to explore these relationships.

In conjunction with understanding the effects of clinically relevant HO on outcomes, it is important to determine possible predictors of clinically relevant HO. The mixed-effects logistic regression, using ORs, identified male sex and higher preoperative VAS scores as significant predictors of clinically relevant HO. Nonsignificant predictors included obesity, endplate coverage, and number of levels treated. Our analysis did not include the use of NSAIDs as a variable in the risk factor analysis, because NSAID use was prohibited by the protocol. Previous studies in arthroplasty, specifically hip arthroplasty, have shown that anti-inflammatory drugs reduce HO. The reason is unclear, but it is hypothesized to be an inhibition of the formation of prostaglandins, which activate local bone growth following trauma.<sup>40</sup> Although hip HO was shown to be reduced with anti-inflammatory use, the mechanism is not fully understood, and the HO etiology appears different between lateral and central regions.<sup>41</sup>

Conducting the first in-depth analysis of HO, we endeavored to understand the various models of analysis and the consistency of findings. The mixed-effects Cox proportional hazard model presents an advantage by estimating hazard ratios, which are direct probability ratios of developing clinically relevant HO of one group compared to another, at any given time. Results of the Cox model were similar to the logistic regression analysis, corroborating significance for male sex and preoperative VAS. Although both models found a correlation between endplate coverage and number of levels treated, only the Cox model produced a strong enough correlation to reach statistical significance. These results pose interesting relationships between HO and predictor variables, but the mechanisms contributing to these relationships are unknown and require further analysis.

Leung et al<sup>16</sup> published results of an observational study of the European Bryan disc. Ninety patients completed 12-month postoperative radiographs, with 17.8% experiencing HO; 6.7% were grade 3 or 4. Leung and his coauthors found that male and older patients were at statistically signif-

icant higher risk of clinically relevant HO. This appears as the first publication to study HO in the cervical spine similarly to previous reports on HO in total joint replacement.<sup>24,25</sup>

Wu et al<sup>26,27</sup> did not grade the severity of HO present, but they found that significantly more HO was present in 2-level cases using the Bryan disc. Because the specific grades of HO were not reported, it is difficult to discern the differences when compared with this analysis. Chang et al<sup>28</sup> analyzed and reported on patients receiving CTDR at C3–C4 versus other cervical levels. Although they found C3–C4 had significantly more and higher grades of HO, the C3–C4 group also had a larger percentage of male patients.<sup>28</sup> Male sex appears in multiple studies as a predictor of higher HO rates, although no study to date has offered an explanation for this.<sup>18,29</sup> Other studies have found that sex did not impact HO. Tu et al<sup>30</sup> found a 1.9% rate of grade 4 HO in 36 patients treated with the Bryan disc, with no differences between sex. However, the average follow-up was less than 24 months and the small patient population would make it difficult to detect differences.<sup>30</sup> Ganbat et al<sup>31</sup> performed 3-dimensional finite element analysis on HO. When subjected to compressive force, the area of the vertebral body uncovered by the implant footprint had the most HO formation. This is consistent with our finding that more implant endplate coverage reduces the odds of HO. Radiographically, this study did not evaluate the presence or severity of preoperative osteophytes, which has also been linked to HO development.<sup>29</sup>

In exploratory analysis we found relationships between sex, BMI, pain, and endplate coverage. These relations stress the importance of analyzing predictors for HO in a multivariate context, as we have done here, because these features potentially confound each other's effect on HO. Although this is the first in-depth analysis of level 1 evidence data on HO, this study has limitations. The analysis was performed posthoc, and the issue could be further studied in prospectively planned studies with preplanned statistical analysis. The findings here are limited for clinical decision-making, because we cannot yet make causal inferences. The rates of HO were shown to progress over time, warranting further research into the relationship of HO and inflammatory response. Because the use of NSAIDs was prohibited by the IDE protocol, the effect on HO rate here is unknown.

There remains a paucity of literature analyzing potential surgical technique and implant-specific causes of HO following CTDR surgery. Further analysis needs to be conducted to understand the significance and relationship between each of these possible predictors, and other potential predictors, such as adjacent-level degeneration, sagittal alignment, and operative levels.

Although spine surgeons have traditionally referred to HO as clinically relevant and nonrelevant, this nomenclature appears to be founded on ROM and not impact on clinical outcomes. Based on this analysis, the largest to date, it seems clear that HO terminology should be more accurately defined as motion-restricting and non-motion-restricting.

## REFERENCES

1. Hilibrand AS, Carlson GD, Palumbo MA, Jones PK, Bohlman HH. Radiculopathy and myelopathy at segments adjacent to the site of a previous anterior cervical arthrodesis. *J Bone Joint Surg Am.* 1999;81(4):519–528.
2. Hsu W. The rationale for cervical disc arthroplasty. *Tech Orthop.* 2010;25(2):84–87.
3. Pracyk JB, Traynelis VC. Treatment of the painful motion segment: cervical arthroplasty. *Spine (Phila Pa 1976).* 2005;30(16 suppl):S23–S32.
4. Robertson JT, Papadopoulos SM, Traynelis VC. Assessment of adjacent-segment disease in patients treated with cervical fusion or arthroplasty: a prospective 2-year study. *J Neurosurg Spine.* 2005;3(6):417–423.
5. Coric D. ISASS policy statement—cervical artificial disc. *Int J Spine Surg.* 2014;8.
6. Coric D, Kim PK, Clemente JD, Boltes MO, Nussbaum M, James S. Prospective randomized study of cervical arthroplasty and anterior cervical discectomy and fusion with long-term follow-up: results in 74 patients from a single site. *J Neurosurg Spine.* 2013;18(1):36–42.
7. Coric D, Nunley PD, Guyer RD, et al. Prospective, randomized, multicenter study of cervical arthroplasty: 269 patients from the Kineflex|C artificial disc investigational device exemption study with a minimum 2-year follow-up: clinical article. *J Neurosurg Spine.* 2011;15(4):348–358.
8. Coric D, Cassis J, Carew JD, Boltes MO. Prospective study of cervical arthroplasty in 98 patients involved in 1 of 3 separate investigational device exemption studies from a single investigational site with a minimum 2-year follow-up: clinical article. *J Neurosurg Spine.* 2010;13(6):715–721.
9. Coric D, Finger F, Boltes P. Prospective randomized controlled study of the Bryan Cervical Disc: early clinical results from a single investigational site. *J Neurosurg Spine.* 2006;4(1):31–35.
10. Zigler JE, Delamarter R, Murrey D, Spivak J, Janssen M. ProDisc-C and anterior cervical discectomy and fusion as surgical treatment for single-level cervical symptomatic degenerative disc disease: five-year results of a Food and Drug Administration study. *Spine (Phila Pa 1976).* 2013;38(3):203–209.

11. Delamarter RB, Murrey D, Janssen ME, et al. Results at 24 months from the prospective, randomized, multicenter Investigational Device Exemption trial of ProDisc-C versus anterior cervical discectomy and fusion with 4-year follow-up and continued access patients. *SAS J*. 2010;4(4):122–128.
12. Hisey MS, Bae HW, Davis RJ, Gaede S, Hoffman G, Kim KD, et al. Prospective, randomized comparison of cervical total disc replacement versus anterior cervical fusion: results at 48 months follow-up. *J Spinal Disord Tech*. 2015;28(4):E237–E243.
13. Hisey MS, Bae HW, Davis R, et al. Multi-center, prospective, randomized, controlled investigational device exemption clinical trial comparing Mobi-C Cervical Artificial Disc to anterior discectomy and fusion in the treatment of symptomatic degenerative disc disease in the cervical spine. *Int J Spine Surg*. 2014;8.
14. Davis RJ, Nunley PD, Kim KD, et al. Two-level total disc replacement with Mobi-C cervical artificial disc versus anterior discectomy and fusion: a prospective, randomized, controlled multicenter clinical trial with 4-year follow-up results. *J Neurosurg Spine*. 2015;22(1):15–25.
15. Davis RJ, Kim KD, Hisey MS, et al. Cervical total disc replacement with the Mobi-C cervical artificial disc compared with anterior discectomy and fusion for treatment of 2-level symptomatic degenerative disc disease: a prospective, randomized, controlled multicenter clinical trial: clinical article. *J Neurosurg Spine*. 2013;19(5):532–545.
16. Leung C, Casey AT, Goffin J, et al. Clinical significance of heterotopic ossification in cervical disc replacement: a prospective multicenter clinical trial. *Neurosurgery*. 2005;57(4):759–763; discussion 763.
17. Yi S, Oh J, Choi G, et al. The fate of heterotopic ossification associated with cervical artificial disc replacement. *Spine (Phila Pa 1976)*. 2014;39(25):2078–2083.
18. Yi S, Shin DA, Kim KN, et al. The predisposing factors for the heterotopic ossification after cervical artificial disc replacement. *Spine J*. 2013;13(9):1048–1054.
19. Yi S, Kim KN, Yang MS, et al. Difference in occurrence of heterotopic ossification according to prosthesis type in the cervical artificial disc replacement. *Spine (Phila Pa 1976)*. 2010;35(16):1556–1561.
20. Zhou HH, Qu Y, Dong RP, Kang MY, Zhao JW. Does heterotopic ossification affect the outcomes of cervical total disc replacement?: a meta-analysis. *Spine (Phila Pa 1976)*. 2015;40(6):E332–E340.
21. Meisel HJ, Jurák L, Antinheimo J, et al. Four-year results of a prospective single-arm study on 200 semi-constrained total cervical disc prostheses: clinical and radiographic outcome. *J Neurosurg Spine*. 2016;25(5):556–565.
22. Mehren C, Suchomel P, Grochulla F, et al. Heterotopic ossification in total cervical artificial disc replacement. *Spine (Phila Pa 1976)*. 2006;31(24):2802–2806.
23. Kang J, Shi C, Gu Y, Yang C, Gao R. Factors that may affect outcome in cervical artificial disc replacement: a systematic review. *Eur Spine J*. 2015;24(9):2023–2032.
24. Ahrengart L, Lindgren U. Heterotopic bone after hip arthroplasty: defining the patient at risk. *Clin Orthop Relat Res*. 1993;293:153–159.
25. Pedersen NW, Kristensen SS, Schmidt SA, Pedersen P, Kjaersgaard-Andersen P. Factors associated with heterotopic bone formation following total hip replacement. *Arch Orthop Trauma Surg*. 1989;108(2):92–95.
26. Wu JC, Huang WC, Tu TH, et al. Differences between soft-disc herniation and spondylosis in cervical arthroplasty: CT-documented heterotopic ossification with minimum 2 years of follow-up. *J Neurosurg Spine*. 2012;16(2):163–171.
27. Wu JC, Huang WC, Tsai TY, et al. Multilevel arthroplasty for cervical spondylosis: more heterotopic ossification at 3 years of follow-up. *Spine (Phila Pa 1976)*. 2012;37(20):E1251–E1259.
28. Chang PY, Chang HK, Wu JC, et al. Differences between C3-4 and other subaxial levels of cervical disc arthroplasty: more heterotopic ossification at the 5-year follow-up. *J Neurosurg Spine*. 2016;24(5):752–759.
29. Jin YJ, Park SB, Kim MJ, Kim KJ, Kim HJ. An analysis of heterotopic ossification in cervical disc arthroplasty: a novel morphologic classification of an ossified mass. *Spine J*. 2013;13(4):408–420.
30. Tu TH, Wu JC, Huang WC, et al. Heterotopic ossification after cervical total disc replacement: determination by CT and effects on clinical outcomes. *J Neurosurg Spine*. 2011;14(4):457–465.
31. Ganbat D, Kim YH, Kim K, Jin YJ, Park WM. Effect of mechanical loading on heterotopic ossification in cervical total disc replacement: a three-dimensional finite element analysis. *Biomech Model Mechanobiol*. 2016;15(5):1191–1199.
32. McAfee PC, Cunningham BW, Devine J, Williams E, Yu-Yahiro J. Classification of heterotopic ossification (HO) in artificial disc replacement. *J Spinal Disord Tech*. 2003;16(4):384–389.
33. Park JH, Rhim SC, Roh SW. Mid-term follow-up of clinical and radiologic outcomes in cervical total disc replacement (Mobi-C): incidence of heterotopic ossification and risk factors. *J Spinal Disord Tech*. 2013;26(3):141–145.
34. Wagner SC, Formby PM, Kang DG, et al. Persistent axial neck pain after cervical disc arthroplasty: a radiographic analysis. *Spine J*. 2016;16(7):851–856.
35. Chen J, Wang X, Bai W, Shen X, Yuan W. Prevalence of heterotopic ossification after cervical total disc arthroplasty: a meta-analysis. *Eur Spine J*. 2012;21(4):674–680.
36. Zhao Y, Zhang Y, Sun Y, Pan S, Zhou F, Liu Z. Application of cervical arthroplasty with Bryan cervical disc: 10-year follow-up results in China. *Spine (Phila Pa 1976)*. 2016;41(2):111–115.
37. Suchomel P, Jurák L, Benes V, Brabec R, Bradác O, Elgawhary S. Clinical results and development of heterotopic ossification in total cervical disc replacement during a 4-year follow-up. *Eur Spine J*. 2010;19(2):307–315.
38. Wu JC, Huang WC, Tsai HW, et al. Differences between 1- and 2-level cervical arthroplasty: more heterotopic ossification in 2-level disc replacement: clinical article. *J Neurosurg Spine*. 2012;16(6):594–600.
39. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007;7(5):541–546.
40. Kraft CT, Agarwal S, Ranganathan K, et al. Trauma-induced heterotopic bone formation and the role of the immune system: a review. *J Trauma Acute Care Surg*. 2016;80(1):156–165.
41. Kjaersgaard-Andersen P, Schmidt SA. Total hip

arthroplasty: the role of antiinflammatory medications in the prevention of heterotopic ossification. *Clin Orthop Relat Res.* 1991(263):78–86.

**Disclosures and COI:** P.D.N. receives royalties and is on the speaker's bureau for LDR Spine (now ZimmerBiomet); K.A.F. received institutional support from LDR Spine (now ZimmerBiomet); K.E.M. is a former employee of LDR Spine (now ZimmerBiomet); and M.B.S. received institutional support from LDR Spine (ZimmerBiomet).

**Corresponding Author:** Pierce D. Nunley, 1500 Line Avenue, Suite 200, Shreveport, LA 71101. Phone: (318) 629-5555; Fax: (318) 629-5556; Email: pnunley@louisianaspine.org

Published 15 August 2018

This manuscript is generously published free of charge by ISASS, the International Society for the Advancement of Spine Surgery. Copyright © 2018 ISASS. To see more or order reprints or permissions, see <http://ijssurgery.com>.