university of groningen

University Medical Center Groningen

University of Groningen

# The Odom Criteria

Broekema, Anne E H; Molenberg, Rob; Kuijlen, Jos M A; Groen, Rob J M; Reneman, Michiel F; Soer, Remko

*Published in:*
Journal of Bone and Joint Surgery: American Volume

*DOI:*
10.2106/JBJS.18.00370

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
2019

Link to publication in University of Groningen/UMCG research database

*Citation for published version (APA):*
Broekema, A. E. H., Molenberg, R., Kuijlen, J. M. A., Groen, R. J. M., Reneman, M. F., & Soer, R. (2019). The Odom Criteria: Validated at Last: A Clinimetric Evaluation in Cervical Spine Surgery. *Journal of Bone and Joint Surgery: American Volume*, *101*(14), 1301-1308. https://doi.org/10.2106/JBJS.18.00370

**Copyright**
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

**Take-down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

Download date: 13-02-2023

Copyright © 2019 by The Journal of Bone and Joint Surgery, Incorporated

# The Odom Criteria: Validated at Last

## A Clinimetric Evaluation in Cervical Spine Surgery

Anne E.H. Broekema, MD, Rob Molenberg, BSc, Jos M.A. Kuijlen, MD, PhD, Rob J.M. Groen, MD, PhD, Michiel F. Reneman, MD, PhD, and Remko Soer, PhD

*Investigation performed at the University of Groningen, University Medical Center Groningen, Department of Neurosurgery, the Netherlands*

**Background:** The Odom criteria, established in 1958, are a widely used, 4-point rating scale for assessing the clinical outcome after cervical spine surgery. Surprisingly, the Odom criteria have never been validated, to our knowledge. The aim of this study was to investigate the reliability and validity of the Odom criteria for the evaluation of surgical procedures of the cervical spine.

**Methods:** Patients with degenerative cervical spine disease were included in the study and divided into 2 subgroups on the basis of their most predominant symptom: myelopathy or radiculopathy. Reliability was assessed with interrater and test-retest design using quadratic weighted kappa coefficients. Construct validity was assessed by means of hypotheses testing. To evaluate whether the Odom criteria could act as a global perceived effect (GPE) scale, we assessed concurrent validity by comparing area under the curve (AUC) values of receiver operating characteristic (ROC) curves for the set of questionnaires.

**Results:** A total of 110 patients were included in the study; 19 were excluded, leaving 91 in our analysis. Reliability assessments showed $\kappa = 0.77$ for overall interrater reliability and $\kappa = 0.93$ for overall test-retest reliability. Interrater reliability was $\kappa = 0.81$ for the radiculopathy subgroup and $\kappa = 0.68$ for the myelopathy subgroup. At least 75% of the hypotheses were met. The AUCs showed similar characteristics between the Odom criteria and GPE scale.

**Conclusions:** The Odom criteria met the predefined criteria for reliability and validity. Therefore, the Odom criteria may be used to assess surgical outcome after a cervical spine procedure, specifically for patients presenting with radicular symptoms. The results of previous studies that have been deemed less trustworthy because of the use of the Odom criteria should be reconsidered.

For the evaluation of surgical procedures, various outcome measures are favored. The Odom criteria[1], since their publication in 1958, have been widely used for assessing general clinical outcome after spine surgery[2-6], particularly cervical spine surgery[7-25]. Surprisingly, the criteria have never been validated, to our knowledge. This limits the interpretation of the numerous studies in which the Odom criteria were used to measure outcomes. In the current guidelines for cervical degenerative disease, the criteria are therefore not recommended as a preferred outcome measure[26-28].

The Odom criteria focus on 2 domains: (1) improvement in preoperative symptoms, and (2) the ability to perform daily activities. These domains indirectly represent items from the World Health Organization (WHO) International Classification of Functioning, Disability and Health (ICF) domains of function, activities, and participation[29]. The surgeon rates the outcome of the patient after surgery on a simple 4-point rating scale, from "poor" to "excellent." Therefore, the Odom criteria provide a brief, generic measure of the outcome following a surgical procedure.

In the existing literature, the Odom criteria have been applied with a great variety in terms of their wording and format[9,10,18,20-25]. There are reports in which only 1 component of the criteria was used or in which the patient was classified by filling out questionnaires[17,19,23]. Even in the original article of Odom et al.[1], "daily activities" were defined differently for each item and were referred to as "physical activities," "work," and "daily occupations."

The aim of the current study was to investigate the reliability and validity of the standardized Odom criteria for the evaluation of surgical procedures of the cervical spine. In order to do so, we assessed the criteria according to the following clinimetric aspects: interrater reliability, test-retest reliability, construct validity, and concurrent validity.

**Disclosure:** The authors indicated that no external funding was received for any aspect of this work. The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (http://links.lww.com/JBJS/F358).

**TABLE I Generalized Odom Criteria**

| Score | Rating | Description |
|---|---|---|
| 1 | Excellent | No symptoms related to cervical disease.<br>Able to perform daily activities without limitations |
| 2 | Good | Moderate symptoms related to cervical disease.<br>Able to perform daily activities without significant limitations |
| 3 | Satisfactory | Slight improvement in symptoms related to cervical disease.<br>Significant limitations in daily activities |
| 4 | Poor | No improvement in, or aggravation of, symptoms related to cervical disease.<br>Not able to perform daily activities |

## Materials and Methods

For this study, the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were used[30]. Participants were assessed at the outpatient clinic on the basis of the Odom criteria at 6 weeks postoperatively (T1) by the neurosurgeon who performed the surgery. During the same visit, they were separately assessed by an independent interviewer (resident or senior medical student). The surgeons (n = 9) and interviewers (n = 4) were not instructed on how to assess the Odom criteria, to make the results applicable to general practice.

The independent interviewer reassessed the patient using the Odom criteria 1 week later (T2) by telephone. The interviewers were blinded to each other's, and previous, scoring results. Participants were asked if their health status had improved or worsened between T1 and T2. Patients whose health status had changed were not included in the analysis for test-retest reliability, as a change in health status might have influenced the Odom criteria scoring between the 2 assessments. Patients filled out a set of questionnaires at baseline (preoperatively, T0), at T1, and 26 weeks after surgery (T3).

### *Patients*

Patients admitted to the University Medical Center Groningen (UMCG), the Netherlands, for surgical treatment of degenerative disease of the cervical spine were invited to participate. Patients provided signed informed consent to participate in this study.

Eligible for inclusion were patients 18 to 85 years of age who were scheduled for surgery because of symptoms of radiculopathy or myelopathy. Various etiologies of cervical disease were included. Patients who had previously undergone cervical spine surgery were excluded. The study was approved by the Medical Ethical Committee of the UMCG.

Following the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines for conducting a clinimetric evaluation of outcome measures such as the Odom criteria, we set a sample size of 100 participants, 50 per subgroup[31]. The 2 subgroups were determined on the basis of the patients' most prominent symptom: radiculopathy or myelopathy. With an expected 10% dropout rate, a total of 110 participants were included.

### *Measurements*

#### Odom Criteria

The interviewer classified the patient according to 1 of the categories of the Odom criteria on the basis of the patient's self-reported history. As there are multiple versions of the Odom criteria in the existing literature, we chose to use criteria composed of general definitions, in order to make the criteria applicable to a wide selection of cervical pathologies (Table I).

**TABLE II Hypotheses Regarding the Strength of Relationships Between the Odom Criteria and PROMs (H1-7) and Regarding Known-Groups Validity (H8-9)***

| Hypothesis | Measure | Correlation Coefficient | Score |
|---|---|---|---|
| H1 | GPE | >0.50 | |
| H2 | VAS arm pain | >0.25[55,58,59] | |
| H3 | mJOA | >0.50[53] (myelopathy subgroup) | |
| H4 | NDI | >0.25[48-52, 57] | |
| H5 | VAS neck pain | >0.25[55,58,59] | |
| H6 | EQ-5D-5L | 0.25 to 0.49[57,59] | |
| H7 | WAI | >0.25[54] | |
| H8 | Change scores of PROMs | | Lower for myelopathy compared with radiculopathy |
| H9 | Mean Odom score | | Significantly higher for myelopathy compared with radiculopathy |

*PROM = patient-reported outcome measure, GPE = global perceived effect, VAS = visual analog scale, mJOA = modified Japanese Orthopaedic Association scale, NDI = Neck Disability Index, EQ-5D-5L = EuroQol 5 Dimensions-5 Levels questionnaire, and WAI = Work Ability Index.

**TABLE III PROMs and General Outcome Scores at T0 (Preoperative Baseline) and T1 (6 Weeks Postoperatively)***

| | Radiculopathy | | Myelopathy | | Between Subgroups |
|---|---|---|---|---|---|
| | Mean Score | P Value | Mean Score | P Value | P Value |
| NDI (0% to 100%) | | | | | |
| T0 | 41.0 ± 15.8 | | 27.9 ± 17.8 | | <0.01 |
| T1 | 24.8 ± 12.8 | | 19.5 ± 15.8 | | |
| Change | −16.2 ± 18.0 | <0.01 | −8.4 ± 14.5 | <0.01 | <0.05 |
| VAS neck pain (0 to 100) | | | | | |
| T0 | 48.3 ± 29.9 | | 33.0 ± 31.5 | | <0.05 |
| T1 | 23.4 ± 21.7 | | 19.3 ± 23.5 | | |
| Change | −24.9 ± 33.8 | <0.01 | −13.7 ± 30.9 | <0.01 | 0.14 |
| VAS arm pain (0 to 100) | | | | | |
| T0 | 58.7 ± 23.7 | | 32.8 ± 30.4 | | <0.01 |
| T1 | 19.4 ± 23.7 | | 14.5 ± 19.4 | | |
| Change | −39.3 ± 29.8 | <0.01 | −18.3 ± 24.8 | <0.01 | <0.01 |
| WAI (0 to 10) | | | | | |
| T0 | 3.7 ± 2.6 | | 4.7 ± 2.9 | | 0.15 |
| T1 | 5.3 ± 2.3 | | 5.7 ± 2.8 | | |
| Change | 1.6 ± 3.4 | <0.01 | 1.0 ± 2.7 | <0.05 | 0.48 |
| EQ-5D-5L (−0.446 to 1) | | | | | |
| T0 | 0.58 ± 0.24 | | 0.65 ± 0.25 | | 0.08 |
| T1 | 0.78 ± 0.14 | | 0.77 ± 0.20 | | |
| Change | 0.20 ± 0.24 | <0.01 | 0.12 ± 0.20 | <0.01 | 0.29 |
| EQ-5D-VAS (0 to 100) | | | | | |
| T0 | 58.6 ± 21.6 | | 66.4 ± 21.2 | | 0.11 |
| T1 | 69.0 ± 21.5 | | 77.0 ± 17.7 | | |
| Change | 10.4 ± 30.3 | <0.01 | 10.6 ± 20.1 | <0.01 | 0.66 |
| mJOA (0 to 18) | | | | | |
| T0 | | | 13.6 ± 2.6 | | |
| T1 | | | 14.9 ± 2.6 | | |
| Change | | | 1.3 ± 2.4 | <0.01 | |
| GPE (1 to 7) | | | | | |
| T1 | 2.5 ± 0.9 | | 2.7 ± 0.8 | | 0.21 |
| Odom (1 to 4) | | | | | |
| T1 | 2.0 ± 0.9 | | 2.3 ± 0.7 | | 0.02 |

*The values are given as the mean and standard deviation. PROM = patient-reported outcome measure, NDI = Neck Disability Index, VAS = visual analog scale, WAI = Work Ability Index, EQ-5D-5L = EuroQol 5 Dimensions-5 Levels questionnaire, mJOA = modified Japanese Orthopaedic Association scale, and GPE = global perceived effect.

**Global Perceived Effect**

The global perceived effect (GPE) was measured on a 7-point scale ranging from 1 to 7 (with 1 indicating "completely improved" and 7 indicating "extremely worsened"). The patient was asked, "(How much) are you improved since the start of your treatment?" GPE scales with a 7-point rating are widely accepted as a reliable way to assess a patient's perceived effect of treatment[32].

**Patient-Reported Outcome Measures (PROMs)**

The following set of PROMs was also used: a visual analog scale (VAS) for arm pain and for neck pain, the Neck Disability Index (NDI), the single-item Work Ability Index (WAI), the EuroQol 5 Dimensions-5 Levels (EQ-5D-5L) health-related quality-of-life survey, and the modified Japanese Orthopaedic Association (mJOA) score[33-38]. For each of these PROMs, validated Dutch language versions were used[39-42].

### *Statistical Methods and Interpretation*

**Reliability**

We assessed interrater and test-retest reliability. As the Odom criteria are ordinal, reliability was calculated using quadratic weighted kappa coefficients, each presented with the associated

**TABLE IV Correlation of PROMs with Odom Criteria and GPE Scale at T1***

| | Total Group | | Radiculopathy Subgroup | | Myelopathy Subgroup | |
|---|---|---|---|---|---|---|
| | Odom | GPE | Odom | GPE | Odom | GPE |
| NDI | | | | | | |
| $R_S$ | 0.36 | 0.30 | 0.49 | 0.46 | 0.35 | 0.19 |
| P value | <0.01 | 0.01 | <0.01 | <0.01 | 0.03 | 0.25 |
| VAS neck pain | | | | | | |
| $R_S$ | 0.17 | 0.18 | 0.34 | 0.42 | 0.08 | 0.02 |
| P value | 0.11 | 0.12 | 0.02 | 0.01 | 0.64 | 0.92 |
| VAS arm pain | | | | | | |
| $R_S$ | 0.37 | 0.24 | 0.53 | 0.35 | 0.31 | 0.21 |
| P value | <0.01 | 0.03 | <0.01 | 0.03 | 0.06 | 0.21 |
| WAI | | | | | | |
| $R_S$ | −0.42 | −0.34 | −0.44 | −0.41 | −0.49 | −0.29 |
| P value | <0.01 | <0.01 | <0.01 | 0.01 | <0.01 | 0.06 |
| EQ-5D-5L | | | | | | |
| $R_S$ | −0.35 | −0.27 | −0.37 | −0.34 | −0.37 | −0.21 |
| P value | <0.01 | 0.02 | 0.01 | 0.04 | 0.02 | 0.19 |
| mJOA | | | | | | |
| $R_S$ | −0.55 | −0.43 | | | −0.55 | −0.43 |
| P value | <0.01 | 0.01 | | | <0.01 | 0.01 |
| GPE | | | | | | |
| $R_S$ | 0.49 | | 0.49 | | 0.46 | |
| P value | <0.01 | | <0.01 | | <0.01 | |

*Spearman correlation coefficients, with the values interpreted as follows: 0.00 to 0.25 = little or no correlation; 0.26 to 0.50 = fair; 0.51 to 0.75 = moderate to good; and >0.75 = good to excellent correlation. A negative sign indicates an inverse correlation. PROM = patient-reported outcome measure, GPE = global perceived effect, NDI = Neck Disability Index, VAS = visual analog scale, WAI = Work Ability Index, EQ-5D-5L = EuroQol 5 Dimensions-5 Levels questionnaire, and mJOA = modified Japanese Orthopaedic Association scale.

standard error (SE)[43,44]. Kappa (κ) scores were interpreted as follows: ≤0.00 = poor reliability, 0.01 to 0.20 = slight, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = substantial, and 0.81 to 1.00 = almost perfect[45].

### Construct Validity

Construct validity was analyzed by means of hypothesis testing. Correlations were calculated using Spearman correlation coefficients, with the interpretation of the correlation coefficients as follows: 0.00 to 0.25 = little or no correlation, 0.26 to 0.50 = fair, 0.51 to 0.75 = moderate to good, and >0.75 = good to excellent[46]. If 75% of the hypotheses are met, construct validity is considered sufficient[47]. Because we assumed the Odom criteria to be a general measure for surgical outcome, it was hypothesized that the Odom scores would have at least a moderate (>0.50) correlation with the GPE scores. It was expected that the relationships of the Odom scores with PROMs would be similar to the known relationships of GPE scales with other PROMs[48-59]. The hypothesized correlations are presented in Table II.

### Known-Groups Validity

The 2 subgroups have different clinical profiles: patients with myelopathy are often operated on to prevent progression of the myelopathy, whereas patients presenting with radiculopathy are operated on for clear improvement of their symptoms, especially improvement of (arm) pain. Therefore, we hypothesized that the postoperative change score (difference from the preoperative value) on the questionnaires would be lower for the myelopathy subgroup. Furthermore, we hypothesized that the myelopathy subgroup would have a significantly higher mean Odom score (indicating a less successful outcome) compared with the radiculopathy subgroup (Table II).

### Concurrent Validity

We investigated whether the Odom criteria could act as a GPE scale. For this purpose, the criteria were dichotomized as "successful" (Odom criteria score of 1 or 2) and "unsuccessful" (Odom criteria score of 3 or 4). Change scores were calculated for the PROMs. Receiver operating characteristic (ROC) curves were plotted, and the area under the curve (AUC) values were calculated. An AUC of >0.70 was considered sufficiently responsive[60]. We compared the AUCs of the Odom criteria with the AUCs of the GPE scale. For the GPE scale, we considered ratings of 1 ("completely improved") and 2 ("much improved") as "successful."

## Results

### *Baseline Characteristics*

Signed informed consent was provided by 110 patients. We excluded 14 patients from our analysis because of missing data or because the patient had undergone another surgical procedure within 6 weeks following the initial surgery. An additional 5 patients had symptoms of both radiculopathy and myelopathy and were therefore also excluded. The remaining 91 patients were included in the analysis. Of these, 49 patients had radiculopathy as the most prominent symptom and 42 patients had myelopathy as the most prominent symptom. The patients in the myelopathy subgroup were significantly older than those in the radiculopathy subgroup (mean [and standard deviation], 60.7 ± 12.7 compared with 52.2 ± 9.8 years; $p < 0.01$). With respect to sex, body mass index (BMI), and American Society of Anesthesiologists (ASA) classification of physical status, the 2 subgroups did not differ significantly.

### *Reliability*

#### Interrater Reliability

Postoperative Odom scores at T1 were obtained for 87 patients. Interrater reliability was as follows: for the total group, κ = 0.77 (SE = 0.12); for the radiculopathy subgroup, κ = 0.81 (SE = 0.11); and for the myelopathy subgroup, κ = 0.68 (SE = 0.30).

#### Test-Retest Reliability

Odom scores at T1 and T2 were available for 88 patients. Of these patients, 13 indicated that their health status had changed and were excluded from analysis. For the remaining patients, test-retest reliability was as follows: for the total group, κ = 0.93 (SE = 0.06); for the radiculopathy subgroup, κ = 0.98 (SE = 0.05); and for the myelopathy subgroup, κ = 0.87 (SE = 0.16).

**TABLE V Results for the Hypotheses at T1***

| | Measure | Hypothesized Correlation/Score | Hypotheses Met at T1 | | |
|---|---|---|---|---|---|
| | | | Total Group | Radiculopathy Subgroup | Myelopathy Subgroup |
| Hypothesis | | | | | |
| H1 | GPE | >0.50 | − | − | − |
| H2 | VAS arm pain | >0.25 | + | + | + |
| H3 | mJOA | >0.50 | NA | NA | + |
| H4 | NDI | >0.25 | + | + | + |
| H5 | VAS neck pain | >0.25 | − | + | − |
| H6 | EQ-5D-5L | 0.25 to 0.49 | + | + | + |
| H7 | WAI | >0.25 | + | + | + |
| H8 | Change scores of PROMs | Lower for myelopathy subgroup | + | + | + |
| H9 | Mean Odom score | Significantly higher for myelopathy subgroup | + | + | + |
| % of hypotheses met | | | 75% | 88% | 78% |

*GPE = global perceived effect, − = hypothesis rejected, VAS = visual analog scale, + = hypothesis met, mJOA = modified Japanese Orthopaedic Association scale, NA = not applicable NDI = Neck Disability Index, EQ-5D-5L = EuroQol 5 Dimensions-5 Levels questionnaire, and WAI = Work Ability Index.

## *Validity*

### Known-Groups Validity

Mean subgroup scores for the preoperative and postoperative questionnaires are presented in Table III. Both subgroups showed a significant improvement in all PROMs after surgery. A larger improvement in scores was detected for the radiculopathy group; however, the difference was only significantly different for the NDI (−16.2 compared with −8.4; $p < 0.05$), the VAS for arm pain (−39.3 compared with −18.3; $p < 0.01$), and the Odom score at 6 weeks (2.0 compared with 2.3; $p = 0.02$). There was no significant difference in mean GPE scores between the subgroups ($p = 0.21$).

### Construct Validity

Table IV presents the correlation coefficients for the PROMs relative to the Odom and GPE scores at T1. As shown in Table V, almost all relationships met the hypotheses. The relationship between the Odom scores and GPE did not meet the hypothesis, with a correlation coefficient of 0.49 noted for the total group. The strength of the correlation between the Odom and GPE scores increased for all groups at T3; in the total group, increasing from r = 0.49 at T1 to r = 0.69 at T3; and in the myelopathy group, from r = 0.46 to r = 0.56. The largest improvement was seen in the radiculopathy group, in which the strength of the correlation increased from r = 0.49 to r = 0.77.

### Concurrent Validity

To analyze whether the Odom criteria could act as a GPE scale, we compared the responsiveness of the questionnaires with the GPE scale as an anchor and with the Odom criteria as an anchor. Table VI shows similar AUCs between the GPE scale

**TABLE VI Comparison of AUCs with Odom Criteria and with GPE Scale as Anchor***

| | Total Group | | Radiculopathy | | Myelopathy | |
|---|---|---|---|---|---|---|
| | Odom | GPE | Odom | GPE | Odom | GPE |
| VAS arm pain | 0.67 | 0.72 | 0.73 | 0.77 | 0.67 | 0.71 |
| VAS neck pain | 0.60 | 0.60 | 0.73 | 0.65 | 0.42 | 0.57 |
| NDI | 0.71 | 0.76 | 0.75 | 0.80 | 0.63 | 0.74 |
| WAI | 0.78 | 0.74 | 0.84 | 0.86 | 0.81 | 0.66 |
| EQ-5D-5L | 0.61 | 0.66 | 0.73 | 0.76 | 0.46 | 0.55 |

*An area under the curve (AUC) of >0.70 was considered sufficient for adequate responsiveness. GPE = global perceived effect, VAS = visual analog scale, NDI = Neck Disability Index, WAI = Work Ability Index, and EQ-5D-5L = EuroQol 5 Dimensions-5 Levels questionnaire.

and the Odom criteria for the questionnaires. AUCs were sufficient for the total group for the NDI and the WAI for both anchors, and for the VAS for arm pain with the GPE scale as an anchor. For the radiculopathy subgroup, all AUCs were sufficient, with the exception of the VAS for neck pain with the GPE scale as an anchor.

## Discussion

### *Interpretation of Results*

The Odom criteria are an outcome measure often considered suboptimal because their clinimetric properties, to our knowledge, have never been studied. Our findings in the current study showed that the Odom criteria had substantial interrater reliability ($\kappa = 0.77$) and an almost perfect test-retest reliability ($\kappa = 0.93$) for patients after cervical spine surgery. This study also provided evidence to support their validity.

The interrater reliability was higher for the radiculopathy subgroup than for the myelopathy subgroup ($\kappa = 0.81$ compared with $\kappa = 0.68$). This could be explained by the different treatment goals of the subgroups. For patients with myelopathy, the goal of surgery is to prevent symptoms from progressing. This causes difficulties for the interviewer classifying the "success" of a surgical procedure. For instance, if the symptoms of a patient with myelopathy are slightly improved postoperatively and the patient has no limitations in daily activities, most surgeons would classify the outcome as "excellent." But literally taken, because not all symptoms are improved, the patient should be classified as "good." We chose not to instruct the investigators in rating the Odom criteria, to make the results applicable to general practice. If we had instructed the investigators, the interrater reliability might have been higher.

The correlation between the Odom criteria and the GPE at T1 was lower than hypothesized ($r = 0.49$). This indicates that the criteria measure a construct or constructs other than the GPE. The GPE scale rates the general improvement of a patient's symptoms. The Odom criteria measure the ability to perform daily activities as well. One explanation could be that this domain is more dominant in the Odom criteria, especially 6 weeks after surgery.

The single-item WAI showed a negative correlation with the Odom criteria of 0.42, which was, after the GPE, the strongest correlation for the total group[54]. This finding supports our hypothesis that the ability to perform daily activities, such as work, is reflected more dominantly in the Odom criteria 6 weeks after surgery than the improvement of symptoms.

Furthermore, some patients will not have returned to work 6 weeks after the surgical procedure, which can influence the Odom criteria but not the GPE. This is supported by the finding that the correlation between the Odom criteria and the GPE after 6 months (T3) was substantially higher ($r = 0.69$) than at 6 weeks (T1).

Similar AUCs were found for the responsiveness of the questionnaires with the Odom criteria as an anchor compared with the GPE scale as an anchor. The Odom criteria have previously been used as an anchor[61]. In that previous report, it appeared that the Oswestry Disability Index, the Million Index, and the VAS for neck pain correlated well with the dichotomized Odom criteria, consistent with our results for the radiculopathy group, for which all of the AUCs were >0.70. This indicates that the Odom criteria can sufficiently differentiate between clinically improved and not improved states in patients with radicular symptoms.

### *Strengths and Limitations*

This is the first study that we are aware of to evaluate the clinimetric aspects of the Odom criteria. The study was designed in accordance with the COSMIN guidelines. Our results are applicable to general clinical practice, with a study population consisting of patients who underwent various surgical procedures of the cervical spine.

A limitation of this study was the relatively high number of participants who were excluded or lost to follow-up (19 of 110 patients). Five of these patients were excluded because they had both radicular and myelopathic symptoms and therefore could not be divided into the distinct subgroups. Furthermore, the second examiners were medical students or residents known to the neurosurgeon and were not independent neurosurgeons. This could have led to different classification styles or some influence on the trainee classification style. Although senior neurosurgeons seem to classify patients more optimistically than do junior surgeons[62], in the present study, differences in scoring were not significant between the examiners. Another limitation of this study was that multiple versions of the Odom criteria exist in the current literature. By generalizing the criteria, we aimed to assess the version that is most applicable to a variety of pathologies and the historical versions of the criteria.

### *Consideration of the Odom Criteria in Clinical and Scientific Practice*

This study provides evidence that the Odom criteria are a reliable and valid method for assessing general success after a cervical spine procedure. A drawback of the Odom criteria is that the criteria aim to measure 2 constructs (improvement in preoperative symptoms and the ability to perform daily activities) using 1 scale. Therefore, the content or the clinical meaning of the Odom score is difficult to interpret. However, the Odom criteria demonstrated results similar to those of a GPE scale as indicated by AUCs and correlations with other questionnaires and, therefore, could be considered as a scale for general surgical success, specifically for patients with radiculopathy.

Furthermore, it is also questionable what exactly defines a "surgical success." For instance, patient satisfaction after spinal surgery does not correspond well with questionnaires regarding pain, disability, and quality of life[57,58,63]. Other studies have used composite scales to define surgical success that are quite elaborate and may not fit use in daily clinical practice[64,65]. The great advantage of the Odom criteria is the simplicity of the 4-point rating scale while indirectly representing

items from the WHO ICF domains of function, activities, and participation.

The validation of the Odom criteria, as presented in this report, puts the results of a number of previous studies in a different perspective. For instance, according to the guidelines for cervical radiculopathy of the North American Spine Society and the Dutch Association of Neurosurgeons, some studies were downgraded in their "level of evidence" partly because the use of the Odom criteria as a nonvalidated outcome measure. On the basis of the results of our present validation, this decision should be reconsidered[8,10,22,26,28]

In summary, the assumption that the Odom criteria are a suboptimal outcome measure should be reconsidered. Our findings demonstrated acceptable results for validity and reliability. The criteria have the advantage of being a simple 4-point rating scale and being easy to use in daily practice for assessing the general clinical outcome after a cervical spine procedure. Although the design and execution of this study were based on current standards and the results are convincing, this is the first study, to our knowledge, in which the clinimetric aspects of the Odom criteria were assessed. Therefore, independent replication of this study is needed.

### *Conclusions*

The Odom criteria meet the predefined criteria for reliability and validity and may be used to measure surgical outcome after a cervical spine procedure, specifically for patients presenting with radicular symptoms. ■

NOTE: The authors thank Diane Steenks and Asmara Pattipeiluhu for their assistance in the data collection and administration of the study.

Anne E.H. Broekema, MD[1]
Rob Molenberg, BSc[1]
Jos M.A. Kuijlen, MD, PhD[1]
Rob J.M. Groen, MD, PhD[1]
Michiel F. Reneman, MD, PhD[2]
Remko Soer, PhD[3,4]

[1]University of Groningen, University Medical Center Groningen, Department of Neurosurgery, the Netherlands

[2]University of Groningen, University Medical Center Groningen, Department of Rehabilitation, the Netherlands

[3]University of Groningen, University Medical Center Groningen, Groningen Spine Center, the Netherlands

[4]Saxion University of Applied Sciences, Enschede, the Netherlands

E-mail address for J.M.A. Kuijlen: j.m.a.kuijlen@umcg.nl

ORCID iD for A.E.H. Broekema: 0000-0002-2657-7540
ORCID iD for R. Molenberg: 0000-0002-6621-2028
ORCID iD for J.M.A. Kuijlen: 0000-0002-2907-5853
ORCID iD for R.J.M. Groen: 0000-0003-3067-763X
ORCID iD for M.F. Reneman: 0000-0002-4058-3437
ORCID iD for R. Soer: 0000-0003-4154-1711

## References

**1.** Odom GL, Finney W, Woodhall B. Cervical disk lesions. J Am Med Assoc. 1958 Jan 4;166(1):23-8.

**2.** Ahmed H, Siam AE, Gouda-Mohamed GM, Boehm H. Surgical treatment of sacroiliac joint infection. J Orthop Traumatol. 2013 Jun;14(2):121-9. Epub 2013 Apr 5.

**3.** Bayerl SH, Finger T, Heiden P, Esfahani-Bayerl N, Topar C, Prinz V, Woitzik J, Dengler J, Vajkoczy P. Radiofrequency denervation for treatment of sacroiliac joint pain-comparison of two different ablation techniques. Neurosurg Rev. 2018 Jul 31. Epub 2018 Jul 31.

**4.** Gibson JN, Depreitere B, Pflugmacher R, Schnake KJ, Fielding LC, Alamin TF, Goffin J. Decompression and paraspinous tension band: a novel treatment method for patients with lumbar spinal stenosis and degenerative spondylolisthesis. Spine J. 2015 Mar 2;15(3)(Suppl):S23-32. Epub 2015 Jan 8.

**5.** Simpson JM, Silveri CP, Balderston RA, Simeone FA, An HS. The results of operations on the lumbar spine in patients who have diabetes mellitus. J Bone Joint Surg Am. 1993 Dec;75(12):1823-9.

**6.** Yang SC, Tsai TT, Chen HS, Fang CJ, Kao YH, Tu YK. Comparison of sacroplasty with or without balloon assistance for the treatment of sacral insufficiency fractures. J Orthop Surg (Hong Kong). 2018 May-Aug;26(2):2309499018782575.

**7.** Park YK, Moon HJ, Kwon TH, Kim JH. Long-term outcomes following anterior foraminotomy for one- or two-level cervical radiculopathy. Eur Spine J. 2013 Jul; 22(7):1489-96. Epub 2013 Feb 16.

**8.** Fessler RG, Khoo LT. Minimally invasive cervical microendoscopic foraminotomy: an initial clinical experience. Neurosurgery. 2002 Nov;51(5)(Suppl):S37-45.

**9.** Palma L, Mariottini A, Carangelo B, Muzii VF, Zalaffi A. Favourable long-term clinical outcome after anterior cervical discectomy. A study on a series of 125 patients undergoing surgery a mean of 11 years earlier. Acta Neurochir (Wien). 2010 Jul;152(7):1145-52.

**10.** Bärlocher CB, Barth A, Krauss JK, Binggeli R, Seiler RW. Comparative evaluation of microdiscectomy only, autograft fusion, polymethylmethacrylate interposition, and threaded titanium cage fusion for treatment of single-level cervical disc disease: a prospective randomized study in 125 patients. Neurosurg Focus. 2002 Jan 15;12(1):E4.

**11.** Fernández-Fairen M, Sala P, Dufoo M Jr, Ballester J, Murcia A, Merzthal L. Anterior cervical fusion with tantalum implant: a prospective randomized controlled study. Spine (Phila Pa 1976). 2008 Mar 1;33(5):465-72.

**12.** Hacker RJ. A randomized prospective study of an anterior cervical interbody fusion device with a minimum of 2 years of follow-up results. J Neurosurg. 2000 Oct; 93(2)(Suppl):222-6.

**13.** Hauerberg J, Kosteljanetz M, Bøge-Rasmussen T, Dons K, Gideon P, Springborg JB, Wagner A. Anterior cervical discectomy with or without fusion with ray titanium cage: a prospective randomized clinical study. Spine (Phila Pa 1976). 2008 Mar 1; 33(5):458-64.

**14.** Kim KT, Kim YB. Comparison between open procedure and tubular retractor assisted procedure for cervical radiculopathy: results of a randomized controlled study. J Korean Med Sci. 2009 Aug;24(4):649-53. Epub 2009 Jul 29.

**15.** Schröder J, Grosse-Dresselhaus F, Schul C, Wassmann H. PMMA versus titanium cage after anterior cervical discectomy - a prospective randomized trial. Zentralbl Neurochir. 2007 Feb;68(1):2-7.

**16.** Thomé C, Krauss JK, Zevgaridis D. A prospective clinical comparison of rectangular titanium cages and iliac crest autografts in anterior cervical discectomy and fusion. Neurosurg Rev. 2004 Jan;27(1):34-41. Epub 2003 Aug 2.

**17.** Shi R, Li J, Liu H, Ding C, Hu T, Li T, Gong Q. Clinical comparison of 2 implantation systems for single-level cervical disk replacement. Orthopedics. 2014 Feb; 37(2):e161-8.

**18.** Quan GM, Vital JM, Hansen S, Pointillart V. Eight-year clinical and radiological follow-up of the Bryan cervical disc arthroplasty. Spine (Phila Pa 1976). 2011 Apr 15; 36(8):639-46.

**19.** Jödicke A, Daentzer D, Kästner S, Asamoto S, Böker DK. Risk factors for outcome and complications of dorsal foraminotomy in cervical disc herniation. Surg Neurol. 2003 Aug;60(2):124-9, discussion :129-30.

**20.** Luo H, Wang J, Liang W, Sheng W. Anterior cervical decompression and fusion versus anterior cervical disc replacement for single-level cervical spondylosis: evaluation of mid-term outcomes. Chin J Tissue Eng Res. 2014;18(48): 7719-25.

**21.** Korinth MC, Krüger A, Oertel MF, Gilsbach JM. Posterior foraminotomy or anterior discectomy with polymethyl methacrylate interbody stabilization for cervical soft disc disease: results in 292 patients with monoradiculopathy. Spine (Phila Pa 1976). 2006 May 15;31(11):1207-14, discussion :1215-6.

**22.** Herkowitz HN, Kurz LT, Overholt DP. Surgical management of cervical soft disc herniation. A comparison between the anterior and posterior approach. Spine (Phila Pa 1976). 1990 Oct;15(10):1026-30.

**23.** Hamburger C, Festenberg FV, Uhl E. Ventral discectomy with pmma interbody fusion for cervical disc disease: long-term results in 249 patients. Spine (Phila Pa 1976). 2001 Feb 1;26(3):249-55.

**24.** O'Shaughnessy BA, Liu JC, Hsieh PC, Koski TR, Ganju A, Ondra SL. Surgical treatment of fixed cervical kyphosis with myelopathy. Spine (Phila Pa 1976). 2008 Apr 1;33(7):771-8.

**25.** Yang X, Song YM, Liu LM, Lü CL, Kong QQ, Tu CQ. [Anterior decompression and fusion with n-HA/PA66 cage for the treatment of lower cervical fracture and dislocation]. Zhongguo Gu Shang. 2014 Feb;27(2):92-6.

**26.** Bono CM, Ghiselli G, Gilbert TJ, Kreiner S, Reitman C, Summers J, Baisden J, Easa J, Fernand R, Lamer T, Matz P, Mazanec D, Resnick DKO. Shaffer W, Sharma A, Timmons R, Toton J. NASS clinical guidelines – diagnosis and treatment of cervical radiculopathy from degenerative disorders. 2010. https://www.spine.org/documents/researchclinicalcare/guidelines/cervicalradiculopathy.pdf. Accessed 2019 Mar 28.

**27.** Congress of Neurological Surgeons. Guideline for the surgical management of cervical degenerative disease. 2009. https://www.cns.org/guidelines/guideline-surgical-management-cervical-degenerative-disease. Accessed 2019 Apr 30.

**28.** Arts M, Bartels R, Bouma GJ, Donk R, Verbeek A, Verhagen W. [Nederlandse Vereniging voor Neurochirurgie. Richtlijn behandeling van cervicaal radiculair syndroom ten gevolge van een cervicale hernia nuclei pulposi]. 2010. https://richtlijnendatabase.nl/richtlijn/cervicaal_radiculair_syndroom_tgv_chnp/cervicaal_radiculair_syndroom_-_startpagina.html. Accessed 2019 Mar 28. Dutch.

**29.** World Health Organization. International Classification of Functioning, Disability and Health: ICF. 2001. http://apps.who.int/iris/bitstream/handle/10665/42407/9241545429.pdf. Accessed 2019 Mar 28.

**30.** Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011 Jan;64(1):96-106. Epub 2010 Jun 17.

**31.** Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49. Epub 2010 Feb 19.

**32.** Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-70.

**33.** El Fassi M, Bocquet V, Majery N, Lair ML, Couffignal S, Mairiaux P. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. BMC Public Health. 2013 Apr 8;13:305.

**34.** Huskisson EC. Measurement of pain. Lancet. 1974 Nov 9;2(7889):1127-31.

**35.** EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199-208.

**36.** Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther. 1991 Sep;14(7):409-15.

**37.** Yonenobu K. Cervical radiculopathy and myelopathy: when and what can surgery contribute to treatment? Eur Spine J. 2000 Feb;9(1):1-7.

**38.** Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013 Sep;22(7):1717-27. Epub 2012 Nov 25.

**39.** Bartels RH, Verbeek AL, Benzel EC, Fehlings MG, Guiot BH. Validation of a translated version of the modified Japanese orthopaedic association score to assess outcomes in cervical spondylotic myelopathy: an approach to globalize outcomes assessment tools. Neurosurgery. 2010 May;66(5):1013-6.

**40.** Soer R, Reneman MF, Speijer BL, Coppes MH, Vroomen PC. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. Spine J. 2012 Nov;12(11):1035-9.

**41.** Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JH, Reneman MF. Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch language versions. Eur Spine J. 2012 Jan;21(1):93-100. Epub 2011 Aug 4.

**42.** de Zwart BC, Frings-Dresen MH, van Duivenbooden JC. Test-retest reliability of the Work Ability Index questionnaire. Occup Med (Lond). 2002 Jun;52(4):177-81.

**43.** Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas. 1973;33:613-9.

**44.** Everitt BS. Moments of the statistics kappa and weighted kappa. Br J Math Stat Psychol. 1968;21(1):97-103.

**45.** Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74.

**46.** Portney LGWA, editor. Foundations of clinical research; applications to practice. 3rd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall; 2009.

**47.** Terwee CB, Bot SDM, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42. Epub 2006 Aug 24.

**48.** Jorritsma W, Dijkstra PU, de Vries GE, Geertzen JH, Reneman MF. Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J. 2012 Dec;21(12):2550-7. Epub 2012 Jul 3.

**49.** Monticone M, Ambrosini E, Vernon H, Brunati R, Rocca B, Foti C, Ferrante S. Responsiveness and minimal important changes for the Neck Disability Index and the Neck Pain Disability Scale in Italian subjects with chronic neck pain. Eur Spine J. 2015 Dec;24(12):2821-7. Epub 2015 Feb 7.

**50.** Takeshita K, Hosono N, Kawaguchi Y, Hasegawa K, Isomura T, Oshima Y, Ono T, Oshina M, Oda T, Kato S, Yonenobu K. Validity, reliability and responsiveness of the Japanese version of the Neck Disability Index. J Orthop Sci. 2013 Jan;18(1):14-21. Epub 2012 Sep 4.

**51.** Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine (Phila Pa 1976). 2006 Mar 1;31(5):598-602.

**52.** Ailliet L, Rubinstein SM, de Vet HC, van Tulder MW, Terwee CB. Reliability, responsiveness and interpretability of the neck disability index-Dutch version in primary care. Eur Spine J. 2015 Jan;24(1):88-93. Epub 2014 May 17.

**53.** Zhou F, Zhang Y, Sun Y, Zhang F, Pan S, Liu Z. Assessment of the minimum clinically important difference in neurological function and quality of life after surgery in cervical spondylotic myelopathy patients: a prospective cohort study. Eur Spine J. 2015 Dec;24(12):2918-23. Epub 2015 Sep 1.

**54.** Leggett S, van der Zee-Neuen A, Boonen A, Beaton DE, Bojinca M, Bosworth A, Dadoun S, Fautrel B, Hagel S, Hofstetter C, Lacaille D, Linton D, Mihai C, Petersson IF, Rogers P, Sergeant JC, Sciré C, Verstappen SM; At-work Productivity Global Measure Working Group. Test-retest reliability and correlations of 5 global measures addressing at-work productivity loss in patients with rheumatic diseases. J Rheumatol. 2016 Feb;43(2):433-9. Epub 2015 Dec 1.

**55.** Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol. 2010 Jul;63(7):760-766.e1. Epub 2010 Jan 8.

**56.** McDonough CM, Tosteson TD, Tosteson AN, Jette AM, Grove MR, Weinstein JN. A longitudinal comparison of 5 preference-weighted health state classification systems in persons with intervertebral disk herniation. Med Decis Making. 2011 Mar-Apr;31(2):270-80. Epub 2010 Nov 22.

**57.** Abtahi AM, Lyman KS, Brodke DS, Lawrence BD, Zhang C, Spiker WR. Patient satisfaction is not associated with self-reported disability in a spine patient population. Clin Spine Surg. 2017 Oct;30(8):E1165-8.

**58.** Hamilton DK, Kong C, Hiratzka J, Contag AG, Ailon T, Line B, Daniels A, Smith JS, Passias P, Protopsaltis T, Sciubba D, Burton D, Shaffrey C, Klineberg E, Mundis G, Kim HJ, Lafage V, Lafage R, Scheer J, Boachie-Adjei O, Bess S, Hart RA. Patient satisfaction after adult spinal deformity surgery does not strongly correlate with health-related quality of life scores, radiographic parameters, or occurrence of complications. Spine (Phila Pa 1976). 2017 May 15;42(10):764-9.

**59.** Parai C, Hägg O, Lind B, Brisby H. The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. Eur Spine J. 2018 Mar;27(3):554-63. Epub 2017 Oct 20.

**60.** de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine, a practical guide. Cambridge: Cambridge University Press; 2011.

**61.** Zoëga B, Kärrholm J, Lind B. Outcome scores in degenerative cervical disc surgery. Eur Spine J. 2000 Apr;9(2):137-43.

**62.** Lattig F, Grob D, Kleinstueck FS, Porchet F, Jeszenszky D, Bartanusz V, O'Riordan D, Mannion AF. Ratings of global outcome at the first post-operative assessment after spinal surgery: how often do the surgeon and patient agree? Eur Spine J. 2009 Aug;18(Suppl 3):386-94. Epub 2009 May 22.

**63.** Roitberg BZ, Thaci B, Auffinger B, Kaplan L, Shen J, Brown FD, Lam S. Comparison between patient and surgeon perception of degenerative spine disease outcomes—a prospective blinded database study. Acta Neurochir (Wien). 2013 May;155(5):757-64. Epub 2013 Mar 7.

**64.** Davis RJ, Kim KD, Hisey MS, Hoffman GA, Bae HW, Gaede SE, Rashbaum RF, Nunley PD, Peterson DL, Stokes JK. 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 Nov;19(5):532-45. Epub 2013 Sep 6.

**65.** Sasso RC, Smucker JD, Hacker RJ, Heller JG. Artificial disc versus fusion: a prospective, randomized study with 2-year follow-up on 99 patients. Spine (Phila Pa 1976). 2007 Dec 15;32(26):2933-40, discussion :2941-2.