



## Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

### Brief Measure Information

**NQF #:** 2643

**Corresponding Measures:**

**Measure Title:** Average change in functional status following lumbar spine fusion surgery

**Measure Steward:** MN Community Measurement

**sp.02. Brief Description of Measure:** For patients age 18 and older undergoing lumbar spine fusion surgery, the average change from pre-operative functional status to one year (nine to fifteen months) post-operative functional status using the Oswestry Disability Index (ODI version 2.1a) patient reported outcome tool.

**1b.01. Developer Rationale:** Lumbar spine surgery, an effective procedure for many spine conditions, may be controversial and less successful for some patients, particularly those with degenerative disc disease. Utilization data indicate up to a fifteen fold increase in the number of complex fusion procedures performed for Medicare beneficiaries (Trends, major medical complications and charges associated with surgery for lumbar spinal stenosis in adults Deyo, RA JAMA April 2010). News articles convey the experiences of some patients who have an increase in intensity of pain and loss of function after surgery. (Back surgery may backfire on patients in pain- NBC News Oct 2010, Doctors getting rich with fusion surgery debunked by studies- BusinessWeek Jan 2011, Pushing back on back surgery- StarTribune Aug 2009)

This PRO measure was developed with a focus on functional status from a patient's perspective to address and understand current gaps in care for patients undergoing lumbar fusion surgery. Other new measures currently included in federal programs assess the ability to administer PRO tools pre and post-operatively, but no measures exist for this population or attempt to reflect the change in score demonstrating the functional status outcome that could be expected for patients undergoing this procedure.

---

**sp.12. Numerator Statement:** There is not a traditional numerator for this measure; the measure is calculating the average change in functional status score from pre-operative to post-operative functional status score. The measure is NOT aiming for a numerator target value for a post-operative ODI score.

For example:

The average change in low back function was an increase in 17.2 points one year post-operatively on a 100 point scale.

**sp.14. Denominator Statement:** Adult patients age and older (no upper age limit) who undergo a lumbar spine fusion procedure during a calendar year performance period (e.g. dates of procedure occurring between 1/1/2013

and 12/31/2013) AND have a completed pre-operative and post-operative ODI patient reported outcome assessments.

**sp.16. Denominator Exclusions:** Exclusions are for patients with spine related cancer, fracture and infection and idiopathic or congenital scoliosis.

---

**Measure Type:** Outcome: PRO-PM

**sp.28. Data Source:**

Paper Medical Records

Instrument-Based Data

Other

**sp.07. Level of Analysis:**

Clinician: Group/Practice

---

**IF Endorsement Maintenance – Original Endorsement Date:** 2015-07-07 04:29 PM

**Most Recent Endorsement Date:** 7/7/2015 4:29:15 PM

---

**IF this measure is included in a composite, NQF Composite#/title:**

**IF this measure is paired/grouped, NQF#/title:**

**sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:**

## 1. Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

**1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.**

[Response Begins]

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

**2021 Submission:**

Updated evidence information here.

**2018 Submission:**

Evidence from the previous submission here.

**1a.01. Provide a logic model.**

*Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.*

[Response Begins]

[Response Ends]

**1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.**

*Describe how and from whom input was obtained.*

[Response Begins]

[Response Ends]

**1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.**

[Response Begins]

[Response Ends]

**1b.01. Briefly explain the rationale for this measure.**

*Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.*

**[Response Begins]**

Lumbar spine surgery, an effective procedure for many spine conditions, may be controversial and less successful for some patients, particularly those with degenerative disc disease. Utilization data indicate up to a fifteen fold increase in the number of complex fusion procedures performed for Medicare beneficiaries (Trends, major medical complications and charges associated with surgery for lumbar spinal stenosis in adults Deyo, RA JAMA April 2010). News articles convey the experiences of some patients who have an increase in intensity of pain and loss of function after surgery. (Back surgery may backfire on patients in pain- NBC News Oct 2010, Doctors getting rich with fusion surgery debunked by studies- BusinessWeek Jan 2011, Pushing back on back surgery- StarTribune Aug 2009)

This PRO measure was developed with a focus on functional status from a patient's perspective to address and understand current gaps in care for patients undergoing lumbar fusion surgery. Other new measures currently included in federal programs assess the ability to administer PRO tools pre and post-operatively, but no measures exist for this population or attempt to reflect the change in score demonstrating the functional status outcome that could be expected for patients undergoing this procedure.

**[Response Ends]**

**1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.**

*Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

**[Response Begins]**

Pilot testing of measures was completed in May of 2014. The measure development work group recommended this measure for state-wide implementation and this recommendation was approved by our MNMCM Measurement and Reporting Committee (MARC) and Board of Directors.

Four practices participated in the pilot testing of this specialty (orthopedic and neurosurgery) measure. Two are orthopedic spine specialty groups, one is a neurosurgical group and one is a multi-specialty practice/ health care system. Pilot data submitted represented 25 surgeons (9 neuro, 16 ortho). Dates of procedures were 4/1/2012 to 9/30/2012 with lag time of 15 months to allow for post-operative follow-up for all patients included in the initial patient population (March 2014). The number of patients included in the initial patient population was 716.

The average functional status change is based on the score of the Oswestry Disability Index (ODI) which is measured on a 100 point scale reflecting percent disability where 0 to 20% is minimal disability, 21 to 40% is moderate disability, 41 to 60% is severe disability, 61 to 80% is crippled and 81 to 100% is bed-bound as described by the tool developer.

For the calculation of this measure, which is accomplished by subtracting the post-op score from the pre-op score, a higher the number of points indicates more improvement in function. Please note that the calculation formula does take into account patients whose scores worsen post-operatively and this is incorporated into the average.

Practice Average Points Increase

Practice A 14.6 points

Practice B 19.5 points

Practice C 18.5 points  
Practice D 10.7 points  
Overall Pilot 17.2 points

Pilot demonstrated rates of pre-op ODI administration at an average of 77.0% (range 61 to 87%) and post-operative one year ODI administration rates at 58.4% (range of 41 to 69%). The rate of having both pre-operative and post-operative ODI was 47.6%. These were fairly good results for pilot testing of a new measure and it is anticipated that tool administration rates will continue to improve as the measure matures into its implementation phase.

Other information about patients included in the pilot:

- \* Average age was 58; 34% of the patients were age 65 and older
- \* 81% white, 4% non-white and 93% English as preferred language
- \* Ability to capture BMI (height and weight) 92%; specialty practices less experienced capturing these values.
- \* 19% incidence of tobacco use with significant variation between the practices (13 to 29%)
- \* History of prior back surgery noted for 36% of the patients
- \* Reasons for procedure: spinal stenosis (40%), spondylolithesis (27%), degenerative disc disease (27%), disc herniation (6%)
- \* BMI distribution: 1%- Underweight, 17%- Healthy weight, 31%- Grade 1 Overweight, 41%- Grade 2 Obese, 10%- Grade 3 Morbid Obesity

**[Response Ends]**

**1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.**

**[Response Begins]**

MOS Short form 36 and Oswestry Disability Index Outcomes in lumbar Fusion: a multicenter experience.

METHODS: The patient population included 270 females and 227 males, with a mean age of 47 years. Sixty-five percent (N=324) had one level fusions and 35% (N=173) had two level fusions. Demographic data collected included age, gender, BMI, surgical history, smoking history and work status.

RESULTS: SF-36 Physical Composite Score (PCS) improved a mean 9.9 points at one year postop and 9.5 points at two years post-op. ODI improved a mean 22.2 points at one year post-op and 22.1 points at two years post-op. SF-36 PCS data for surgical approach subgroups revealed greater improvement in the ALIF group as compared to the PSF, PLIF/TLIF, or 360\_ fusion groups (12.6 points vs. 8.8, 9.3, 8.4 points) at 1 year post-op. At 2 years post op, there was greater improvement (p<.02) in the ALIF and PSF groups as compared to the PLIF/TLIF and 360\_ fusion groups (13.8 and 11.2 points vs. 7.7 and 6.3 points). SF-36 PCS data demonstrated similar baseline scores for patients with and without prior decompression, but a significantly greater rate of improvement (11.3 vs. 7.2 points, p<.002) for patients without prior lumbar decompression surgery. The ODI data indicated a significantly greater disability at baseline in the prior decompression group, with greater improvement (21.7 vs. 17.5 points) in patients without prior surgery.

CONCLUSIONS: This study documents improved outcomes, based on SF-36 and ODI scores, in patients undergoing lumbar fusion for one and two level degenerative disc disease. The findings also demonstrate efficacy for all of the surgical techniques studied, suggesting that surgeons can appropriately select the surgical strategy with which they are most adept.

MOS Short form 36 and Oswestry Disability Index Outcomes in lumbar Fusion: a multicenter experience. Glassman et al. The Spine Journal 6 2006.

SPORT (The Spine Patient Outcomes Research Trial)

Surgical versus Nonsurgical Treatment for Lumbar Degenerative Spondylolisthesis

BACKGROUND: Management of degenerative spondylolisthesis with spinal stenosis is controversial.

Surgery is widely used, but its effectiveness in comparison with that of nonsurgical treatment has not been

demonstrated in controlled trials.

**METHODS:** Surgical candidates from 13 centers in 11 U.S. states who had at least 12 weeks of symptoms and image-confirmed degenerative spondylolisthesis were offered enrollment in a randomized cohort or an observational cohort. Treatment was standard decompressive laminectomy (with or without fusion) or usual nonsurgical care.

The primary outcome measures were the Medical Outcomes Study 36-Item Short-Form General Health Survey (SF-36) bodily pain and physical function scores (100-point scales, with higher scores indicating less severe symptoms) and the modified Oswestry Disability Index (100-point scale, with lower scores indicating less severe symptoms) at 6 weeks, 3 months, 6 months, 1 year, and 2 years.

**RESULTS:** We enrolled 304 patients in the randomized cohort and 303 in the observational cohort. The baseline characteristics of the two cohorts were similar. The one-year crossover rates were high in the randomized cohort (approximately 40% in each direction) but moderate in the observational cohort (17% crossover to surgery and 3% crossover to nonsurgical care). The intention-to-treat analysis for the randomized cohort showed no statistically significant effects for the primary outcomes. The as-treated analysis for both cohorts combined showed a significant advantage for surgery at 3 months that increased at 1 year and diminished only slightly at 2 years. The treatment effects at 2 years were 18.1 for bodily pain (95% confidence interval [CI], 14.5 to 21.7), 18.3 for physical function (95% CI, 14.6 to 21.9), and -16.7 for the Oswestry Disability Index (95% CI, -19.5 to -13.9). There was little evidence of harm from either treatment.

**CONCLUSIONS:** In nonrandomized as-treated comparisons with careful control for potentially confounding baseline factors, patients with degenerative spondylolisthesis and spinal stenosis treated surgically showed substantially greater improvement in pain and function during a period of 2 years than patients treated nonsurgically. (ClinicalTrials.gov number, NCT00000409.) Surgical versus Nonsurgical Treatment for Lumbar Degenerative Spondylolisthesis Weinstein et al N Engl J Med 2007;356:2257-70

**[Response Ends]**

**1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.**

*Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

**[Response Begins]**

Demographics of pilot patients:

N = 716

57% female; 43% male

Average age was 58; 34% of the patients were age 65 and older

81% white, 4% non-white and 93% English as preferred language

19% incidence of tobacco use with significant variation between the practices (13 to 29%)

History of prior back surgery noted for 36% of the patients

BMI distribution:

1% Underweight

17% Healthy weight

31% Grade 1 Overweight

41% Grade 2 Obese

10% Grade 3 Morbid Obesity

**[Response Ends]**

**1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.**

**[Response Begins]**

**[Response Ends]**

## 2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

**spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.**

[Response Begins]

[Response Ends]

**spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.**

**For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.**

*For example, specifications may have been updated based on suggestions from a previous NQF CDP review.*

[Response Begins]

[Response Ends]

**sp.01. Provide the measure title.**

*Measure titles should be concise yet convey who and what is being measured (see [What Good Looks Like](#)).*

[Response Begins]

Average change in functional status following lumbar spine fusion surgery

[Response Ends]

**sp.02. Provide a brief description of the measure.**

*Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).*

[Response Begins]

For patients age 18 and older undergoing lumbar spine fusion surgery, the average change from pre-operative functional status to one year (nine to fifteen months) post-operative functional status using the Oswestry Disability Index (ODI version 2.1a) patient reported outcome tool.

[Response Ends]

**sp.04. Check all the clinical condition/topic areas that apply to your measure, below.**

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*



- *Surgery: General*

**[Response Begins]**

Musculoskeletal

Musculoskeletal: Low Back Pain

Surgery

**[Response Ends]**

**sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.**

**[Response Begins]**

Health and Functional Status: Change

Person-and Family-Centered Care: Person-and Family-Centered Care

**[Response Ends]**

**sp.06. Select one or more target population categories.**

*Select only those target populations which can be stratified in the reporting of the measure's result.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

**[Response Ends]**

**sp.07. Select the levels of analysis that apply to your measure.**

*Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Clinician: Group/Practice

**[Response Ends]**

**sp.08. Indicate the care settings that apply to your measure.**

*Check ONLY the settings for which the measure is SPECIFIED and TESTED.*

**[Response Begins]**

Outpatient Services

**[Response Ends]**

**sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.**

*Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".*

**[Response Begins]**

New measure; specs to be publicly reported on our corporate website Feb 2015 at <http://mncm.org/submitting-data/training-and-guidance/> but all field specifications and definitions included in data dictionary S.2b.

**[Response Ends]**

**sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

*Attach an excel or csv file; if this poses an issue, [contact staff](#). Provide descriptors for any codes. Use one file with multiple worksheets, if needed.*

**[Response Begins]**

No data dictionary/code table – all information provided in the submission form

**[Response Ends]**

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.12. State the numerator.**

*Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).*

*DO NOT include the rationale for the measure.*

**[Response Begins]**

There is not a traditional numerator for this measure; the measure is calculating the average change in functional status score from pre-operative to post-operative functional status score. The measure is NOT aiming for a numerator target value for a post-operative ODI score.

For example:

The average change in low back function was an increase in 17.2 points one year post-operatively on a 100 point scale.

**[Response Ends]**

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.13. Provide details needed to calculate the numerator.**

*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

There is not a traditional numerator for this measure; the measure is calculating the average change in functional status score from pre-operative to post-operative functional status score. The measure is NOT aiming for a numerator target value for a post-operative ODI score.

The average change is calculated as follows:

Change is first calculated for each patient and then changed scores are summed and then an average is determined. Measure calculation takes into account those patients that have an improvement and those patients whose function decreases post-operatively. Example below:

Patient Pre-op ODI Post-op ODI Change in ODI

Patient A 47 18 29

Patient B 45 52 -7

Patient C 56 12 44

Patient D 62 25 37

Patient E 42 57 -15

Patient F 51 10 41

Patient G 62 25 37

Patient H 43 20 23

Patient I 74 35 39

Patient J 59 23 36

Average change in ODI one year post-op 26.4 points on a 100 point scale

**[Response Ends]**

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.14. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

Adult patients age and older (no upper age limit) who undergo a lumbar spine fusion procedure during a calendar year performance period (e.g. dates of procedure occurring between 1/1/2013 and 12/31/2013) AND have a completed pre-operative and post-operative ODI patient reported outcome assessments.

**[Response Ends]**

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.15. Provide details needed to calculate the denominator.**

*All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

The initial patient population is adult patients age 18 and older (no upper age limit) who undergo a lumbar spine fusion procedure during a calendar year performance period (e.g. dates of procedure occurring between 1/1/2013 and 12/31/2013).

CPT procedure codes: 22533, 22534, 22558, 22586 22612, 22630, and 22633.

If any portion of the lumbar spine is fused (L1 to L5), the patient is to be included. If the fusion of the lumbar spine also incorporates thoracic vertebrae, the patient is to be included.

Inclusion in the denominator that measures the average change between pre-operative and post-operative functional status requires completion of a patient reported outcome assessment tool (ODI) BOTH pre-operatively (within three months prior to the procedure) AND one year post-operatively (nine to fifteen months after the procedure)

The denominator for calculating the average change in function at a practice level is those patients included in the initial patient population who have both a completed pre-operative and post-operative Oswestry Disability Index patient reported outcome tool (ODI version 2.1a)

**[Response Ends]**

**sp.16. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

Exclusions are for patients with spine related cancer, fracture and infection and idiopathic or congenital scoliosis.

**[Response Ends]**

**sp.17. Provide details needed to calculate the denominator exclusions.**

*All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Patients who are undergoing a lumbar spine fusion procedure for an acute fracture (trauma), metastatic or bone cancer, infection or scoliosis are not included in this patient population because their expected course of care and outcomes could be significantly different from the population of patients undergoing the procedure for relief of

back and/or leg pain (degenerative disc disease, disc herniation, stenosis or spondylolisthesis). ICD-9/ ICD-10 diagnosis codes for exclusions are provided in the data dictionary at S.2.b

**[Response Ends]**

**sp.18. Provide all information required to stratify the measure results, if necessary.**

*Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.*

**[Response Begins]**

Clinical Condition Reason for Procedure field is collected for purposes of stratification (potential) or use in a risk adjustment model (more likely). The choices for this variable are: 1 = Degenerative Disc Disease, 2 = Disc Herniation, 3 = Spinal Stenosis, 4 = Spondylolisthesis. These conditions are definable by ICD-9/ ICD-10 codes and are provided in the data dictionary at S.2.b.

The use of this variable for stratification of outcomes is dependent on procedure volume at the practice level; it has been our experience so far that the volumes at a practice level do not support reliable stratification by four variables as they may result in volumes that do not meet our standards for public reporting at the practice level. These variables, however, are important for several reasons. They may prove appropriate for inclusion in a future risk adjustment model. They also serve analytical purposes for further understanding of the patient reported outcome rates as some of the conditions represent an area of controversy in terms of appropriateness of procedures and successful outcomes for the patient.

**[Response Ends]**

**sp.19. Select the risk adjustment type.**

*Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.*

**[Response Begins]**

Statistical risk model

**[Response Ends]**

**sp.20. Select the most relevant type of score.**

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Continuous variable, e.g. average

**[Response Ends]**

**sp.21. Select the appropriate interpretation of the measure score.**

*Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*

**[Response Begins]**

Better quality = Higher score

**[Response Ends]**

**sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

*Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.*

**[Response Begins]**

Please also refer to measure flow logic in the data dictionary in S.2.b and flow chart in Appendix A-1

Initial patient population:

Was the patient born on or prior to 01/01/xxxx?

Did the patient undergo a lumbar fusion (any portion of the lumbar spine) procedure between 01/01/2013 to 12/31/2013? Patients who had fusion of the lumbar spine which incorporate the thoracic vertebrae are included.

Does the patient have one of the following CPT codes?

22533, 22534, 22558, 22586, 22612, 22630, 22633

Inclusion in Denominator (has pre-op and post-op ODI)

Valid date in the Pre-op ODI Date field? No = remove from denominator; Yes continue

Is the Pre-op ODI Date field within 3 months prior to the procedure? No = remove from denominator; Yes continue

Is there a value in the Pre-op ODI Summary Score field? Yes = Pre-op ODI Hold this score for calculation if postop score is present, if No evaluate if individual responses submitted for score calculation.

Are there at least 8 completed value (valid 0 to 5) responses in the following fields? Pre-op ODI, Pain Pre-op ODI Care, Pre-op ODI Lifting, Pre-op ODI Walking, Pre-op ODI Sitting, Pre-op ODI Standing, Pre-op ODI Sleeping, Pre-op ODI Sex, Pre-op ODI Social, Pre-op ODI Travelling. If Yes = Pre-op ODI Hold this score for calculation if postop score is present, if No remove from the denominator.

Is the 1 Yr Post-op ODI Date field within nine to fifteen months after the Date of Procedure? No = remove from denominator; Yes continue.

Is there a value in the 1 Yr Post-op ODI Summary Score field? If Yes 1 Yr Post-op ODI Hold this score for calculation, if No evaluate if individual responses submitted for score calculation.

Are there at least 8 completed value ( valid 0 to 5) responses in the following fields? 1 Yr Post-op ODI Pain, 1 Yr Post-op Care, 1 Yr Post-op Lifting, 1 Yr Post-op Walking, 1 Yr Post-op Sitting, 1 Yr Post-op Standing, 1 Yr Post-op Sleeping, 1 Yr Post-op ODI Sex, 1 Yr Post-op ODI Social, 1 Yr Post-op ODI Travelling. If Yes = Hold this score for calculation, if No remove from denominator.

For each patient remaining in the denominator calculate the change in function by taking the pre-op ODI score and subtracting the one year post-op ODI score. Save this change score.

To calculate the rate of average change in functional status for the practice; average the change in function score.  
Example:

Patient Pre-op ODI Post-op ODI Change in ODI

Patient A 47 18 29

Patient B 45 52 -7

Patient C 56 12 44

Patient D 62 25 37

Patient E 42 57 -15

Patient F 51 10 41

Patient G 62 25 37

Patient H 43 20 23

Patient I 74 35 39

Patient J 59 23 36

Average change in ODI one year post-op 26.4

**[Response Ends]**

**sp.23. Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.**

**[Response Begins]**

**[Response Ends]**

**sp.24. Indicate the responder for your instrument.**

**[Response Begins]**

**[Response Ends]**

**sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

**[Response Begins]**

This procedurally based measure is based on the full population of eligible patients; sampling is not used.

The patient reported outcome tool for this measure is the Oswestry Disability Index, version 2.1a. We have not had issues related to proxy completion for this adult measure. Developer instructions for the tool (Jeremy Fairbank and MAPI Trust, Inc.) indicate that the tool should be completed by the patient.

**[Response Ends]**

**sp.26. Identify whether and how proxy responses are allowed.**

**[Response Begins]**

**[Response Ends]**

**sp.27. Survey/Patient-reported data.**

*Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.*

**[Response Begins]**

MNCM also calculates rates for tool administration as this measure is dependent on consistent administration of PRO assessment tools to patients. Prior to any use of these outcome measures, we evaluate the rates of:

- \* Rate of pre-operative ODI administered within three months prior to the date of procedure
- \* Rate of post-operative ODI administered within nine to fifteen months post-operatively (one year)
- \* Total and rate of patients who have both pre-operative and post-operative tool administration (denominator)

During the measure development process, the work group anticipated fairly high rates of pre-operative administration and had an expectation that 70% of patients could be captured post-operatively, based on the percentage of patients they were seeing for a one year post-op visit. It is noted that a face-to-face visit is not required for a follow-up ODI assessment; mail or email administration is acceptable.

Pilot demonstrated rates of pre-op ODI administration at an average of 77.0% (range 61 to 87%) and post-operative one year ODI administration rates at 58.4% (range of 41 to 69%). The rate of having both pre-operative and post-operative ODI was 47.6%. These were fairly good results for pilot testing of a new measure and it is anticipated that tool administration rates will continue to improve as the measure matures into its implementation phase.

**[Response Ends]**

**sp.28. Select only the data sources for which the measure is specified.**

**[Response Begins]**

Instrument-Based Data

Paper Medical Records

**[Response Ends]**

**sp.29. Identify the specific data source or data collection instrument.**

*For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.*

**[Response Begins]**

Oswestry Disability Index (ODI) version 2.1a

A ten item self-administered questionnaire with a six point Likert response scale. Items are scored on a 0 to 5 scale with 0 indicating no limitation of function due to pain and 5 indicating major functional disability due to back pain. Time for patient completion is 3 to 5 minutes. Languages available are English and Spanish. The tool is available for use in clinical practice at no cost and can be obtained by completing a user agreement with MAPI Trust, Inc.

The ODI is a valid, reliable, and responsive condition-specific assessment tool that is suited for use in clinical practice. It is easy to administer and score, objectifies client's complaints, and monitors effects of therapy. The ODI shows good construct validity; internal consistency is rated as acceptable; test-retest reliability and responsiveness have been shown to be high; and burden of administration is low. Internal consistency with Cronbach's alpha in ranges from .17 to .87 with test re-test reliability ranges of  $r = 0.83$  to  $0.99$  and intraclass correlation coefficient values from  $0.84$  to  $0.94$ . (Vinarin Psychometric properties and clinical usefulness of the Oswestry Disability Index Journal of Chiropractic Medicine 2008).

**[Response Ends]**

**sp.30. Provide the data collection instrument.**

**[Response Begins]**

**[Response Ends]**

**2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).**

*Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:*

**Current Submission:**

*Updated testing information here.*

**Previous Submission:**

*Testing from the previous submission here.*

**[Response Begins]**

**[Response Ends]**



**2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).**

***Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:***

***Current Submission:***

*Updated testing information here.*

***Previous Submission:***

*Testing from the previous submission here.*

**[Response Begins]**

**[Response Ends]**

**2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?**

**[Response Begins]**

**[Response Ends]**

**2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.**

**Please update the Scientific Acceptability: Validity - Other Threats to Validity section.**

**Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.**

**[Response Begins]**

**[Response Ends]**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.

- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the [2021 Measure Evaluation Criteria and Guidance](#).

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

## Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measure scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of \$25 in cost for an episode of care (e.g., \$5,000 v. \$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

### **2021 Submission:**

Updated testing information here.

### **2018 Submission:**

Testing from the previous submission here.

### **2a.01. Select only the data sources for which the measure is tested.**

**[Response Begins]**

**[Response Ends]**

### **2a.02. If an existing dataset was used, identify the specific dataset.**

*The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).*

**[Response Begins]**

[Response Ends]

**2a.03. Provide the dates of the data used in testing.**

*Use the following format: "MM-DD-YYYY - MM-DD-YYYY"*

[Response Begins]

[Response Ends]

**2a.04. Select the levels of analysis for which the measure is tested.**

*Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

[Response Ends]

**2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).**

*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.*

[Response Begins]

[Response Ends]

**2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.**

*If there is a minimum case count used for testing, that minimum must be reflected in the specifications.*

[Response Begins]

[Response Ends]

**2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.**

[Response Begins]

[Response Ends]

**2a.08. List the social risk factors that were available and analyzed.**

*For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.*

[Response Begins]

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter “see validity testing section of data elements”; and enter “N/A” for 2a.09 and 2a.10.

**2a.09. Select the level of reliability testing conducted.**

*Choose one or both levels.*

[Response Begins]

[Response Ends]

**2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.**

*Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.*

[Response Begins]

[Response Ends]

**2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?**

*For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, [NQF Measure Evaluation Criteria](#)).*

[Response Begins]

[Response Ends]

**2a.12. Interpret the results, in terms of how they demonstrate reliability.**

*(In other words, what do the results mean and what are the norms for the test conducted?)*

[Response Begins]

[Response Ends]

**2b.01. Select the level of validity testing that was conducted.**

[Response Begins]

[Response Ends]

**2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.**

*Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.*

[Response Begins]

[Response Ends]

**2b.03. Provide the statistical results from validity testing.**

*Examples may include correlations or t-test results.*

[Response Begins]

[Response Ends]

**2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)**

[Response Begins]

[Response Ends]

**2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.**

*Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.*

[Response Begins]

[Response Ends]

**2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.**

*Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.*

[Response Begins]

[Response Ends]

**2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.**

*In other words, what do the results mean in terms of statistical and meaningful differences?*

[Response Begins]

[Response Ends]

**2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.**

*Describe the steps—do not just name a method; what statistical analysis was used.*

[Response Begins]

[Response Ends]

**2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.**

*For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).*

[Response Begins]

[Response Ends]

**2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.**

*In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.*

[Response Begins]

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eQCMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

**2b.11. Indicate whether there is more than one set of specifications for this measure.**

[Response Begins]

[Response Ends]

**2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.**

*Describe the steps—do not just name a method. Indicate what statistical analysis was used.*

[Response Begins]

[Response Ends]

**2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.**

*Examples may include correlation, and/or rank order.*

[Response Begins]

[Response Ends]

**2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.**

*In other words, what do the results mean and what are the norms for the test conducted.*

[Response Begins]

[Response Ends]

**2b.15. Indicate whether the measure uses exclusions.**

[Response Begins]

[Response Ends]

**2b.16. Describe the method of testing exclusions and what was tested.**

*Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?*

[Response Begins]

[Response Ends]

**2b.17. Provide the statistical results from testing exclusions.**

*Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.*

[Response Begins]

[Response Ends]



**2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.**

*In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.*

[Response Begins]

[Response Ends]

**2b.19. Check all methods used to address risk factors.**

[Response Begins]

[Response Ends]

**2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.**

[Response Begins]

[Response Ends]

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

[Response Begins]

[Response Ends]

**2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.**

[Response Begins]

[Response Ends]

**2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.**

*Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of  $p < 0.10$  or other statistical tests; correlation of  $x$  or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).*

[Response Begins]

[Response Ends]

**2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.**

[Response Begins]

[Response Ends]

**2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.**

*Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.*

[Response Begins]

[Response Ends]

**2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.**

*Validation testing should be conducted in a data set that is separate from the one used to develop the model.*

[Response Begins]

[Response Ends]

**2b.27. Provide risk model discrimination statistics.**

*For example, provide c-statistics or R-squared values.*

[Response Begins]

[Response Ends]

**2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).**

[Response Begins]

[Response Ends]

**2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.**

*The preferred file format is .png, but most image formats are acceptable.*

[Response Begins]

[Response Ends]

**2b.30. Provide the results of the risk stratification analysis.**

[Response Begins]

[Response Ends]

**2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).**

*In other words, what do the results mean and what are the norms for the test conducted?*

**[Response Begins]**

**[Response Ends]**

**2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.**

*Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.*

**[Response Begins]**

**[Response Ends]**

### 3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

---

**3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.**

**[Response Begins]**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

**[Response Ends]**

**3.02. Detail to what extent the specified data elements are available electronically in defined fields.**

*In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.*

**[Response Begins]**

ALL data elements are in defined fields in electronic health records (EHRs)

**[Response Ends]**

**3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.**

**[Response Begins]**

**[Response Ends]**

**3.04. Describe any efforts to develop an eCQM.**

**[Response Begins]**

**[Response Ends]**

**3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

**[Response Begins]**

**[Response Ends]**

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

**3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),**

**Attach the fee schedule here, if applicable.**

**[Response Begins]**

The PRO tool- Oswestry disability Index is available for free for use in clinical practice by completing a user agreement with MAPI Research Trust Inc., the copyright holder at <http://www.mapi-trust.org/>. MNCM paid a one-time license fee with MAPI to allow the posting of the ODI tool (2.1a) on our password protected data portal to allow practices in MN and bordering communities access to a pdf of the tool without the need for a separate user agreement with MAPI. There are no fees associated with submitting patient level data to the MNCM data portal for rate calculation. There are costs to the practices in the collection and extraction of their data for an annual submission of data.

**[Response Ends]**

## 4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

---

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

### 4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]

Public Reporting

Payment Program

Regulatory and Accreditation Programs

[Response Ends]

### 4a.02. Check all planned uses.

[Response Begins]

Public reporting

Payment Program

[Response Ends]

### 4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

*For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?*

[Response Begins]

Plans in place for publicly reporting in 2015 on our consumer facing website MN HealthScores at [www.mnhealthscores.org/](http://www.mnhealthscores.org/)

[Response Ends]

**4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.**

*A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*

**[Response Begins]**

Planned Use:

\* Included in the MN Department of Health (MDH) Statewide Quality Reporting and Measurement System. Mandatory data collection and reporting under 2008 MN Health Reform Legislation. MNCM was a subcontractor to MDH for measure development exploring the concept of low back pain. Statewide implementation is planned for submission in April/May 2015 for dates of procedure 1/1/2013 to 12/31/2013 with follow-up assessment period through March 31, 2015.

\* Was recommended for inclusion in federal programs by NQF's Measure Application Partnership

\* Was included in the PQRS registry based measures and part of the initial recommendations for CMS Proposed Rule. CMS final rule publication expected in Dec 2014.

**[Response Ends]**

**4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

*Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.*

**[Response Begins]**

**[Response Ends]**

**4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

**[Response Begins]**

**[Response Ends]**

**4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.**

**[Response Begins]**

**[Response Ends]**

**4a.08. Summarize the feedback obtained from those being measured.**

**[Response Begins]**

**[Response Ends]**

**4a.09. Summarize the feedback obtained from other users.**

**[Response Begins]**

**[Response Ends]**

**4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

**[Response Begins]**

**[Response Ends]**

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

**[Response Begins]**

**[Response Ends]**

**4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.**

**[Response Begins]**

No negative unintended consequences identified during pilot. As measure moves forward with wider implementation; need to be mindful of potential gaming (only surveying patients that are doing well); this can be managed through monitoring the pre-operative and post-operative PRO administration rates. One would expect a fairly high pre-operative rate (80's to 90's) that can be incorporated into pre-operative processes and a post-operative capture rate

**[Response Ends]**

**4b.03. Explain any unexpected benefits realized from implementation of this measure.**

**[Response Begins]**

**[Response Ends]**



## 5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

---

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

### 5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

### 5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

### 5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

[Response Ends]

### 5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

No

[Response Ends]

### 5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

Significant differences in these two measures; related but not competing. Only commonality is the desire to measure change in functional status. Target populations, settings of care and provider types are completely different as are the mechanisms for measuring change. # 0425 targets physical therapy settings and providers, for a population of patients with low back pain, and uses a proprietary (monthly/ per provider fee based) web-based CAT tool. Our measure focuses on patients undergoing lumbar fusion procedures, focus on orthopedic and neurosurgery providers in the ambulatory setting (pre and post procedure) and utilizes a valid, free PRO tool with strong psychometric properties that is easy to administer (3 to 5 min to complete) and easy to score.

[Response Ends]

**5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.**

*Provide analyses when possible.*

**[Response Begins]**

Measures do not address the same target population, providers or setting of care. They are related but not competing.

**[Response Ends]**

## Appendix

Supplemental materials may be provided in an appendix.:

## Contact Information

**Measure Steward (Intellectual Property Owner):** MN Community Measurement

**Measure Steward Point of Contact:** Cole, Collette, cole@mncm.org

**Measure Developer if different from Measure Steward:** MN Community Measurement

**Measure Developer Point(s) of Contact:** Cole, Collette, cole@mncm.org

## Additional Information

**1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.**

[Response Begins]

[Response Ends]

**2. List the workgroup/panel members' names and organizations.**

*Describe the members' role in measure development.*

[Response Begins]

Paul Huddleston, MD/ Clinical Provider: Orthopedic Surg; Chair/ Mayo Clinic  
Ann Parr, MD/ Clinical Provider: Neurosurgery/ University of MN Physicians  
Daryll Dykes, MD/ Clinical Provider: Orthopedic Surgery/ Med & Surg Spine Consultants MN  
Denis McCarren/ Clinical Provider: Physical Therapy/ HealthPartners  
Glenn Buttermann, MD/ Clinical Provider: Orthopedic Surgery/ Midwest Spine Institute  
Manuel Pinto, MD/ Clinical Provider: Orthopedic Surgery/ Twin Cities Spine Center  
Michael Goertz, MD/ Clinical Provider: Occupational Medicine/ HealthPartners  
Tom Marr, MD/ Clinical Provider/ Health Plan/ HealthPartners  
Jill Coleman/ Data Analyst/ Essentia Health  
Lyla Westrup/ QI or Clinic Admin/ Twin Cities Spine Center  
Shari Ohland/ QI or Clinic Admin/ Midwest Spine Institute  
Chuck McKinzie, MD/ Health Plan/ MARC Member (2012)/ PrimeWest  
Ann Robinow/ Consumer/ MARC Member  
Chris Norton/ Consumer/ MARC Member  
Kris Soegaard/ Employer/ MARC Member/ MN Health Action Group  
Cally Vinz/ ICSI Low Back Pain/ ICSI Institute Clinical Systems Imp.  
Jeff Schiff, MD MBA/ State Agency/ Department of Human Services  
Collette Piten/ Facilitator/ Measure Development/ MN Community Measurement

[Response Ends]

**3. Indicate the year the measure was first released.**

[Response Begins]

[Response Ends]

**4. Indicate the month and year of the most recent revision.**

[Response Begins]

[Response Ends]

**5. Indicate the frequency of review, or an update schedule, for this measure.**

[Response Begins]

Annual review and update

[Response Ends]

**6. Indicate the next scheduled update or review of this measure.**

**[Response Begins]**

**[Response Ends]**

**7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

© MN Community Measurement, 2014. All rights reserved.  
Oswestry Disability Index Copyright  
ODI © Jeremy Fairbank, 1980. All Rights Reserved.  
ODI - United States/English - Version of 29 Jul 11 - Mapi Institute.  
ID6287 / ODI\_AU2.1a\_eng-US.doc

**[Response Ends]**

**8. State any disclaimers, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

**[Response Ends]**

**9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

**[Response Ends]**