Elsevier Editorial System(tm) for World Neurosurgery

Manuscript Draft

Manuscript Number: WNS-18-824

Title: Can we assess the success of surgery for degenerative spinal diseases by using patients' recall of their preoperative status?

Article Type: Original Article

Keywords: degenerative spine diseases; patient recall; patient-reported outcomes; recall bias; spine surgery

Corresponding Author: Dr. Ricardo Rodrigues, M.B.

Corresponding Author's Institution: Faculty of Medicine, University of Porto, Porto, Portugal

First Author: Ricardo Rodrigues, M.B.

Order of Authors: Ricardo Rodrigues, M.B.; Pedro Santos Silva, MD; Marisa Cunha, MD, MSc; Rui Vaz, MD, PhD; Paulo Pereira, MD, PhD

Abstract: Background:
Patients' recall of their preoperative status is seldom used to assess surgical outcomes because of concerns of inaccuracy and bias.
Objective:
The present study aims to measure the significance of this recall bias and its repercussion on patients' recollection of their preoperative status.
Methods:
Patients submitted to surgery due to degenerative spine diseases during the period of one year (n=198) were included in this study. EQ-5D (including EQ VAS), COMI Neck (including Neck Pain and Shoulder/Arm Pain NRS), COMI Back (including Back Pain and Buttock/Leg Pain NRS), NDI and ODI were completed preoperatively. One year after surgery, patients were asked to complete 2 sets of the same questionnaires, one regarding their postoperative status and the other one regarding their recall of the preoperative status.
Results:
There was poor to moderate agreement between recalled and collected preoperative scores for all PROMs. Patients' recollection of their preoperative status was accurate for patients who underwent cervical spine surgery, but not after lumbar spine surgery. Patients satisfied with the outcome after lumbar spine surgery recalled significantly worse scores compared to the preoperatively collected.
Conclusions:
Using patients' recall of their preoperative status may lead to an overestimation of the surgery effectiveness, particularly for lumbar spine surgery. The self-assessed surgery effectiveness interferes with the recollection of the baseline status.

***Cover Letter**

Click here to download Cover Letter: Cover Letter.docx

25 February 2018

Editorial Department of World Neurosurgery

Dear Editor of World Neurosurgery:

I am pleased to submit an original research article entitled “Can we assess the success of surgery for degenerative spinal diseases by using patients’ recall of their preoperative status?” for consideration for publication in World Neurosurgery. In this manuscript, we show that using patients’ recall of their preoperative status may lead to an overestimation of the surgery effectiveness, particularly for lumbar spine surgery. We also demonstrated that the self-assessed surgery effectiveness interferes with the recollection of the baseline status.

I, Ricardo Rodrigues, certify that this manuscript is a unique submission and is not being considered for publication, in part or in full, with any other source in any medium.

Thank you for your consideration!

Yours sincerely,

Ricardo Rodrigues

Faculty of Medicine,University of Porto

Porto, Portugal

Alameda Prof. Hernâni Monteiro

4200 - 319 Porto, Portugal

mdricardorodrigues@gmail.com

Declarations of Interest

We wish to confirm that there are **no known conflicts of interest** associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Abbreviations list:

CI – Confidence Interval

COMI - Core Outcome Measures Index

EQ-5D - EuroQol Five Dimension Questionnaire

MCID - Minimal clinically important difference

NDI - Neck Disability Index

NRS - Numeric Rating Scale

ODI - Oswestry Disability Index

PROMS - Patient-reported Outcome Measures

SD - Standard Deviation

VAS – Visual Analog Scale

Highlights:

- Cervical patients accurately recalled their preoperative status
- Lumbar patients recalled significantly worse scores than preoperatively collected
- Identification of the main symptom showed a higher agreement than severity of pain
- Effectiveness of surgery influences recall bias
- Collecting data retrospectively may not be accurate, especially in lumbar patients

Introduction

Spinal surgery aims to improve function, provide a better quality of life and relieve pain. Conventional clinical tools are not appropriate to quantify the severity of spine pathology and assess patients' evolution over time, since they do not accurately characterize the patients' subjective perception of the change in their clinical status after surgery comparing to their preoperative condition. Recently, clinicians and researchers have been relying on patient-reported outcome measures (PROMs) to better appraise these data[1]. In response to this growing demand for outcome evaluation, EuroSpine (Spine Society of Europe) and the University of Bern developed, in the year 2000, an international spinal registry – Spine Tango[2].

In the last ten years, there has been an increasing number of new PROMs being developed and employed in spine patients, despite the lack of standardization across different medical centres. They can be divided into two categories: general health assessment tools, and disease-specific outcomes[1]. Among the first category, the EuroQol Five Dimension Questionnaire (EQ-5D) is a standardized and effective PROM, commonly used in spine surgery[3,4]. It was developed in 1990 by the EuroQol group, in order to function as a simple tool that can be used both in clinics and postal surveys. Regarding disease-specific outcomes, two of the most frequently used PROMs are the Oswestry Disability Index (ODI) and the Neck Disability Index (NDI), indicated for back pain and neck pain, respectively[1,5]. Lastly, Spine Tango considers the Core Outcome Measures Index (COMI) as their official tool to assess patient-based outcome, which is validated as a self-assessment form which includes a numeric rating scale (NRS) regarding neck, arm/shoulder, back and leg/buttock pain[6,7].

Despite the usefulness of PROMs, their inherent subjective nature leaves them prone to inaccuracy of self-reporting and patients' different interpretation of the questionnaires. In

addition, when used retrospectively, the concern about a recall bias of their preoperative status is an issue. Recognizing the magnitude of this recall bias and its relation to the different types of pain and preoperative status is fundamental to evaluate the validity of using PROMs in a recall setting.

Recall bias is a systematic error, frequently present in clinical research that involves questionnaires or interviews, whose risk estimate can be biased away from or towards the null[8]. Its effect has been studied for primary health care visits[9], prostate cancer[10], hip[11,12] and knee arthroplasty[13], back pain[14–16], and, more recently, for lumbar decompression and fusion surgery[17]. The objective of this study was to investigate the magnitude of patient recall bias, at least 12 months after the surgery, on general quality of life, cervical and lumbar pain; its influence in the reliability of patients' recollection; and compare data between collected preoperative, recalled preoperative and postoperative status.

Materials and Methods

We conducted an observational longitudinal study in the Neurosurgery Department of Centro Hospitalar São João, which includes all patients submitted to surgery due to a degenerative spine disease during the period of one year.

All patients older than 18 years of age who underwent spine surgery due to degenerative cervical or lumbar pathologies during the year of 2016 were included in this study. Patients who have a valid preoperative, recalled and postoperative score for at least one of the PROMs were included in the statistical analyses. Patients were excluded if the indication for surgery was not a degenerative disease; if their follow-up was lost; if they did not complete the questionnaires; if they deceased or if they did not consent to participate in this study.

Clinical data regarding the patient and surgery were collected from Spine Tango

registry.

Disease-specific and general health outcomes were used. The preoperative questionnaires were completed by patients at the time of hospital admission for surgery as part of the standard clinical practice. Two sets of the same questionnaires were sent to the patients by mail at least 12 months after the surgery, one referring to the actual status and the other to their recall of the preoperative status.

EQ-5D was used to assess the quality of life, based on five measures, which translate to 245 different health states. Patients were questioned about mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each health state has a different weighted value set assigned to it, which is converted to an index score that ranges from -0.536 (worst health status) to 1.0 (perfect health)[18]. EQ-5D also includes a VAS (Health VAS), which ranges from 0 (worst health) to 100 (best health). This questionnaire has been validated for spine surgery[19]. Minimal clinically important difference (MCID) for this questionnaire is 0.30[20].

To measure neck-related outcomes, COMI neck and NDI were selected. COMI is a PROM whose reliability has been extensively studied in spine surgery[21–24]. It comprises a NRS for neck pain and another for arm/shoulder pain, which ranges from 0 (no pain) to 10 (maximum pain). COMI scores are expressed as a percentage, with higher scores representing worse patient's status. MCID value for COMI Neck is 1.7, for Neck Pain NRS is 2.6 and for arm/shoulder Pain NRS is 4.1[24,25]. NDI evaluates neck-related disability based on pain intensity, personal care, lifting, reading, headache, concentration, work, driving, sleeping and recreation. NDI scores are also expressed as a percentage with higher scores meaning higher limitations. MCID for this form is 17.3-percentage points[25].

Concerning lumbar-related outcomes, the PROMs used were COMI Back and ODI. COMI Back is a similar form to COMI Neck, but measuring back pain instead of neck pain

and leg/buttock pain instead of arm/shoulder pain. MCID for COMI Back is 1.7, Back Pain NRS has a MCID of 1.8 points and leg/buttock Pain NRS has a MCID of 1.9 points[24]. Oswestry Disability Index consists of 10 questions, regarding pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and travelling. Its total score is expressed as a percentage, with higher scores representing higher back-related disability. MCID is 12.8-percentage points for ODI[26].

Nonparametric tests were used to account for the non-normal distribution of the different PROMs scores.

Multiple regression analysis was applied to evaluate agreement for each PROM as a function of age, gender, and time between forms.

The weighted kappa statistic was used to measure the degree of agreement for categorical variables between collected preoperative and recalled preoperative scores. In this test, total scores for the different PROMs, VAS and NRS scores were converted into five equally distributed categories. A coefficient (κ_w) of 0 indicates an agreement expected by chance alone. Kappa coefficients of less than 0.2 indicate poor agreement; 0.20 to 0.40 fair agreement; 0.41 to 0.60 moderate agreement; 0.61 to 0.80 good agreement; and >0.80 indicate excellent agreement[27].

The Wilcoxon signed-rank test was applied to determine the median difference between scores (collected preoperative, recalled preoperative, postoperative scores). We also performed subgroup analyses for three clinical groups: cervical – which contains cervical myelopathy and cervical radiculopathy patients; neurogenic claudication; and lumbar radiculopathy. Cervical patients were combined into a single group due to its small sample size.

Recall bias was expressed as a percentage for the different PROMs and also according to the clinical group. If the difference between recalled and collected scores was smaller than

the respective MCID for that form, it was considered as no recall bias. Patients who reported a recall bias greater than the MCID, were sorted into two different groups according to the direction of the bias: worse recall or better recall.

A subgroup analysis based on the effectiveness of surgery was performed comparing recalled with preoperative scores. For this test, patients were sorted into two groups according to the self-rated outcome of surgery, as evaluated by COMI: helpful surgery – which contains "helped a lot", "helped" and "helped only little"; and unhelpful surgery – which contains "didn't help" and "made things worse". We further subdivided the "successful surgery" group into different lumbar clinical groups, but not the "unsuccessful surgery" group due to its small sample size.

Statistical analyses were conducted using IBM SPSS (version 25), statistical significance was accepted at $p<0.05$ level and all p values were two-tailed.

Results

A total of 230 patients were assessed for eligibility, of which 198 were included in the study - the inclusion rate was 86% (Figure 1). There was no significant difference between demographic or clinical characteristics of patients included in the study and of patients excluded from it. The mean age at surgery was 54 years (range 19-85, standard deviation (SD)=13.1), 58.6% were female and the mean time between surgery and recalled PROMs was 14.1 months (range 12–19, SD=1.9) (Table 1). Data about pathology and location are described in Table 2.

There was no association between EQ-5D, COMI Neck, COMI Back, ODI or NDI recall agreement and gender, age, or time between forms.

Weighted kappa statistic showed moderate recall agreement for NDI (0.420, 95% Confidence Interval (CI) 0.221 – 0.620). COMI Back was the only PROM with poor

agreement (0.127, 95% CI 0.030 – 0.224). Fair recall agreement was found for all the other PROMs (Table 3).

Considering the entire cohort, recalled EQ-5D and Health VAS scores are significantly worse that the collected preoperative (Table 4). However, this does not happen in cervical patients, who have similar recalled and collected scores in these PROMs. In addition, cervical patients also have comparable recalled and collected scores for COMI Neck and NRS for neck and arm/shoulder pain. NDI is the only cervical PROM whose recall is significantly worse. On the contrary, for patients who underwent lumbar surgery, all recalled scores are significantly worse than the preoperatively collected. Lastly, patients reported significantly better postoperative median scores for all PROMs and in all clinical categories, than the ones they reported preoperatively.

Recalled versus collected preoperative scores for different PROMs are illustrated in Figure 2, in terms of recall bias and taking into consideration the MCID. In most PROMs, the majority of patients did not present a clinically significant recall bias. Nevertheless, a considerable percentage of lumbar patients reported a worse recall than the preoperatively collected scores, thus, overestimating the effect of surgery. Back pain and disability showed the highest percentage of worse recall bias.

Patients who self-assessed the surgery as helpful, recalled significantly worse scores compared to the preoperatively collected, except for cervical patients whose scores are identical. On the contrary, patients who labelled the surgery as not helpful, recalled similar median scores as preoperatively for all PROMs (Table 5).

Discussion

The present study demonstrated poor to moderate agreement between recalled and collected preoperative scores for all PROMs, suggesting these measures are not appropriate

to use retrospectively. Lingard *et al* also reported similar results for total knee arthroplasty[13]. Identification of the main symptom showed a higher weighted kappa, and thus higher agreement, than severity of pain, a finding which is in line with previous reports[16]. COMI Neck, NDI, neck and arm/shoulder Pain NRS exhibited fair to moderate recall agreement. Previous studies also showed fair to moderate test-retest reliability for NDI and neck NRS[28]. COMI Back, ODI, back and leg/buttock Pain NRS demonstrated poor to fair agreement, suggesting that lumbar patients have a less accurate recall of their preoperative status than cervical patients.

Overall, patients are likely to recall their preoperative quality of life as being more severe than it actually was. The same trend was also seen in knee surgery[11]. However, the present study is the first to document that cervical patients can accurately recall their quality of life, preoperative status and pain, but not disability, one year after the surgery. Lumbar patients' recollection of their preoperative status was significantly worse than they had reported preoperatively. These results are in line with other studies suggesting that retrospectively recalling of preoperative quality of life, pain and disability does not yield the same results as collecting these data preoperatively, specifically for lumbar pain[15,17]. Surgeon's records were also reported not to provide an accurate estimate of outcomes data[29].

When taking into account the MCID, the majority of cervical patients exhibited a difference between recalled and collected preoperative scores which is not considered significant from a clinical perspective, which re-launches the discussion about the validity of using recalled data in clinical studies including these patients. Both neurogenic claudication and lumbar radiculopathy patients showed a substantial worse recall bias, suggesting that data collected retrospectively is not suitable to use in these clinical scenarios.

Patients who labelled the surgery as helpful reported similar recalled scores for cervical surgery and worse recalled scores for lumbar surgery, leading to an overestimation of the

effect of surgery in this last clinical category. Nevertheless, patients who considered the surgery as not helpful, recalled similar scores as the preoperatively collected, hence accurately estimating the effect of surgery. This finding was also reported in a previous study[30], which observed that, despite the standardization of the PROMs used, patients assess themselves differently as a function of the spine surgery's effectiveness, creating a 'moving goal post' and, thus, making more challenging the interpretation of the scores over time.

This study has several limitations. Firstly, the follow-up period considered (at least 12 months after surgery) may have been too long, thus affecting the accuracy of patients' recall. Other studies showed recall was accurate when the interval for recollection was 48 hours[31], one week[32], and 3 months[12]. Despite this, a long time span is essential to evaluate the outcomes of spinal procedures, and most papers reporting surgical outcomes use at least a 12-months follow-up. Hence, assessing the long-term recall bias was explicitly the purpose of this research. Secondly, the number of patients eligible for this study was not big enough to allow for a subgroup analysis by cervical pathology groups. Finally, MCID values display a considerable variation among publications and according to the pathology and type of surgery[26], which limits their applicability to the population of patients in this study and influences the importance of the results obtained to the clinical setting.

Careful selection of patients and questions included in the PROMs used may contribute to a better recall, avoiding the overestimation of the effectiveness of spinal surgery. Recall adjustments were employed in the past[33], but strategies to overcome this bias also have their own limitations[34]. Another conclusion of the present study, which is in line with previous publications[11,17], is that recalled scores are very unlikely to underestimate the effect of surgery.

The results presented herein emphasize the importance of collecting data prospectively and not retrospectively to assess the outcomes of spinal surgery. Good quality data from

registries and cohort studies are paramount, particularly at a time when there are still few prospective randomized studies investigating the efficacy of spinal procedures, due to their financial resources, time and ethical constraints. Furthermore, patients provide unique insights into the effectiveness of spine surgery so, they must be given a central role in reporting and evaluating the different outcomes, making PROMs vital not only in the clinical setting, but also in academic research.

Conclusions

This study shows that relying on patients' recollection of their preoperative status is accurate for patients with cervical degenerative diseases, but not for lumbar degenerative patients, since it may lead to an overestimation of the surgery effectiveness. Furthermore, the self-assessed surgery effectiveness influences the recollection of the baseline status.

References

1. McCormick JD, Werner BC, Shimer AL. Patient-reported Outcome Measures in Spine Surgery. *J Am Acad Orthop Surg*. 2013;21(2):99-107. doi:10.5435/JAAOS-21-02-99
2. Aghayev E, Sobottke R, Munting E, et al. The international Spine Registry SPINE TANGO. https://www.eurospine.org/cm_data/Spine_Tango_Report_International_2015_19_9_16.pdf. Accessed January 28, 2018.
3. Mueller B, Carreon LY, Glassman SD. Comparison of the EuroQOL-5D With the Oswestry Disability Index, Back and Leg Pain Scores in Patients With Degenerative Lumbar Spine Pathology. *Spine (Phila Pa 1976)*. 2013;38(9):757-761. doi:10.1097/BRS.0b013e31827ab803
4. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
5. Guzman JZ, Cutler HS, Connolly J, et al. Patient-reported outcome instruments in spine surgery. *Spine (Phila Pa 1976)*. 2016;41(5):429-437. doi:10.1097/BRS.0000000000001211
6. Ferrer M, Pellisé F, Escudero O, et al. Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine (Phila Pa 1976)*. 2006;31(12):1372-1379. doi:10.1097/01.brs.0000218477.53318.bc
7. Mannion AF, Elfering A, Staerkle R, et al. Outcome assessment in low back pain: how low can you go? *Eur Spine J*. 2005;14:1014-1026. doi:10.1007/s00586-005-0911-9
8. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol*. 1990;43(1):87-91. doi:10.1016/0895-4356(90)90060-3
9. Brusco NK, Watts JJ. Empirical evidence of recall bias for primary health care visits. *BMC Health Serv Res*. 2015;15:381. doi:10.1186/s12913-015-1039-1
10. Litwin MS, McGuigan KA. Accuracy of recall in health-related quality-of-life assessment among men treated for prostate cancer. *J Clin Oncol*. 1999;17(9):2882-2888. doi:10.1200/JCO.1999.17.9.2882
11. Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Med Care*. 1995;33(4 Suppl):AS77-88. doi:10.2307/3766614
12. Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS. A comparison between patient recall and concurrent measurement of preoperative quality of life outcome in total hip arthroplasty. *J*

Arthroplasty. 2008;23(6):843-849. doi:10.1016/j.arth.2007.07.020
13. Lingard EA, Wright EA, Sledge CB. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Jt Surg - Ser A*. 2001;83(8):1149-1156. doi:10.2106/00004623-200108000-00003
14. Serbic D, Pincus T. Diagnostic uncertainty and recall bias in chronic low back pain. *Pain*. 2014;155(8):1540-1546. doi:10.1016/j.pain.2014.04.030
15. Pain B, Vidal X, Herna A, et al. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine (Phila Pa 1976)*. 2005;30(3):365-368. doi:10.1097/01.brs.0000152096.48237.7c
16. Dawson EG, Kanim LEA, Sra P, et al. Low back pain recollection versus concurrent accounts: Outcomes analysis. *Spine (Phila Pa 1976)*. 2002;27(9):984-993. doi:10.1097/00007632-200205010-00020
17. Aleem IS, Duncan J, Ahmed AM, et al. Do lumbar decompression and fusion patients recall their preoperative status?: A cohort study of recall bias in patient-reported outcomes. *Spine (Phila Pa 1976)*. 2017;42(2):128-134. doi:10.1097/BRS.0000000000001682
18. Ferreira LN, Ferreira PL, Pereira LN, Oppe M. The valuation of the EQ-5D in Portugal. *Qual Life Res*. 2014;23(2):413-423. doi:10.1007/s11136-013-0448-z
19. Solberg TK, Olsen J-A, Ingebrigtsen T, Hofoss D, Nygaard OP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. *Eur Spine J*. 2005;14(10):1000-1007. doi:10.1007/s00586-005-0898-2
20. Solberg T, Johnsen LG, Nygaard ØP, Grotle M. Can we define success criteria for lumbar disc surgery? Estimates for a substantial amount of improvement in core outcome measures. *Acta Orthop*. 2013;84(2):196-201. doi:10.3109/17453674.2013.786634
21. Fankhauser CD, Mutter U, Aghayev E, Mannion AF. Validity and responsiveness of the Core Outcome Measures Index (COMI) for the neck. *Eur Spine J*. 2012;21(1):101-114. doi:10.1007/s00586-011-1921-4
22. Monticone M, Ferrante S, Maggioni S, et al. Reliability, validity and responsiveness of the cross-culturally adapted Italian version of the core outcome measures index (COMI) for the neck. *Eur Spine J*. 2014;23(4):863-872. doi:10.1007/s00586-013-3092-y
23. Miekisiak G, Banach M, Kiwic G, et al. Reliability and validity of the Polish version of the Core Outcome Measures Index for the neck. *Eur Spine J*. 2014;23(4):898-903. doi:10.1007/s00586-013-3129-2
24. Damasceno LHF, Rocha PAG, Barbosa ES, et al. Cross-cultural

adaptation and assessment of the reliability and validity of the Core Outcome Measures Index (COMI) for the Brazilian-Portuguese language. *Eur Spine J*. 2012;21(7):1273-1282. doi:10.1007/s00586-011-2100-3

25. Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion. *J Neurosurg Spine*. 2013;18(2):154-160. doi:10.3171/2012.10.SPINE12312
26. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and Pain Scales. *Spine J*. 2008;8(6):968-974. doi:10.1016/j.spinee.2007.11.006
27. Altman DG. *Practical Statistics for Medical Research*. Chapman and Hall; 1991. https://books.google.pt/books/about/Practical_Statistics_for_Medical_Researc.html?id=v-walRnRxWQC&redir_esc=y. Accessed January 31, 2018.
28. Cleland JA, Childs JD, Whitman JM. Psychometric Properties of the Neck Disability Index and Numeric Pain Rating Scale in Patients With Mechanical Neck Pain. *Arch Phys Med Rehabil*. 2008;89(1):69-74. doi:10.1016/j.apmr.2007.08.126
29. Edwards CC, Karpitskaya Y, Cha C, et al. Accurate identification of adverse outcomes after cervical spine surgery. *J Bone Joint Surg Am*. 2004;86-A(2):251-256. http://www.ncbi.nlm.nih.gov/pubmed/14960668.
30. Schwartz CE, Sajobi TT, Lix LM, Quaranto BR, Finkelstein JA. Changing values, changing outcomes: The influence of reprioritization response shift on outcome assessment after spine surgery. *Qual Life Res*. 2013;22(9):2255-2264. doi:10.1007/s11136-013-0377-x
31. Babul N, Darke AC, Johnson DH, Charron-Vincent K. Using memory for pain in analgesic research. *Ann Pharmacother*. 1993;27(1):9-12. doi:10.1177/106002809302700101
32. Singer AJ, Kowalska A, Thode J. Ability of patients to accurately recall the severity of acute painful events. *Acad Emerg Med*. 2001;8(3):292-295. doi:10.1111/j.1553-2712.2001.tb01310.x
33. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. *Health Qual Life Outcomes*. 2010;8:65. doi:10.1186/1477-7525-8-65

34. Raphael K. Recall bias: A proposal for assessment and control. *Int J Epidemiol*. 1987;16(2):167-170. doi:10.1093/ije/16.2.167

Table(s)

Click here to download Table(s): Tables.docx

TABLE 1. Clinical Characteristics of Included Patients		
Female, %		58.6
Age at surgery, years, mean (SD)		54 (13.1)
Age at surgery, years, range		19 - 85
Previous spine surgeries, %	None	80.9
	One	12.9
	Two	5.7
	Three	0.5
Previous spine surgeries at same level, %	None	87.4
	One	11.1
	Two	1.5
Previous treatments for main pathology, %	None	6.1
	Surgical	1.5
	< 3 months conservative	8.7
	3 – 6 months conservative	14.0
	6 – 12 months conservative	14.4
	>12 months conservative	55.3
Time between forms, months, mean (SD)		14.1 (1.9)
Time between forms, months, range		12 - 19

SD, standard deviation.

TABLE 2. Pathology and Location

Level of Intervention, %	Upper cervical	1.0
	Mid lower cervical	27.4
	Thoraco-lumbar	1.0
	Lumbar	52.7
	Lumbo-sacral	17.9
Clinical Group, %	Cervical myelopathy	22.3
	Cervical radiculopathy	6.1
	Neurogenic claudication	30.5
	Lumbar radiculopathy	41.1
Type of Degeneration, %	Disc herniation/bulging	75.4
	Central stenosis	35.0
	Lateral stenosis	19.3
	Foraminal stenosis	9.1
	Degenerative disc disease	31.0
	Degenerative deformity	2.0
	Degenerative spondylolisthesis	13.2
	Myelopathy	15.2
	Facet joint arthrosis	6.1
	Other	0.5

TABLE 3. Agreement and Correlation Between Collected and Recalled Preoperative Scores

TABLE 4. Comparison of Recalled Preoperative and Postoperative to Collected Preoperative Median Scores by Clinical Category

	EQ-5D	Health VAS	COMI Neck	Neck Pain NRS	Arm/Shoulder Pain NRS	NDI	COMI Back	Back Pain NRS	Leg/Buttock Pain NRS	ODI

	K_w	95% CI
EQ-5D	**0.240***	**0.149-0.332**
Mobility	0.289*	0.157-0.421
Self-Care	0.267*	0.162-0.372
Usual Activities	0.319*	0.200-0.437
Pain/Discomfort	0.144*	0.041-0.246
Anxiety/Depression	0.282*	0.170-0.393
Health VAS	0.083	-0.003-0.168
COMI Neck	**0.297***	**0.074-0.519**
Main Symptom	0.464*	0.247-0.682
Neck Pain NRS	0.424*	0.218-9.629
Arm/Shoulder Pain NRS	0.382*	0.179-0.584
Neck-related Function	0.188*	-0.024-0.399
Well-being	0.332*	0.112-0.552
Quality of Life	0.177	-0.040-0.395
Disability	0.260*	0.028-0.492
Disability (social role)	0.424*	0-217-0.631
COMI Back	**0.127***	**0.030-0.224**
Main Symptom	0.422*	0.287-0.557
Back Pain NRS	0.131*	0.009-0.253
Leg/Buttock Pain NRS	0.143*	0.016-0.261
Back-related Function	0.150*	0.055-0.245
Well-being	0.175*	0.014-0.336
Quality of Life	0.198*	0.070-0.326
Disability	0.253*	0.119-0.386
Disability (social role)	0.314*	0.180-0.449
NDI	**0.420***	**0.221-0.620**
Pain Intensity	0.237*	0.056-0.417
Personal Care	0.194*	0.006-0.382
Lifting	0.263*	0.081-0.445
Reading	0.385*	0.201-0.569
Headache	0.509*	0.343-0.674
Concentration	0.382*	0.182-0.582
Work	0.422*	0.218-0.627
Driving	0.295*	00.005-0.586
Sleeping	0.353*	0.183-0.543
Recreation	0.280*	0.065-0.494
ODI	**0.271***	**0.170-0.371**
Pain Intensity	0.065	-0.010-0.140
Personal Care	0.396*	0.293-0.499
Walking	0.409*	0.295-0.523
Lifting	0.262*	0.141-0.384
Sitting	0.203*	0.098-0.308
Standing	0.291*	0.183-0.399
Sleeping	0.192*	0.076-0.308
Sex Life	0.448*	0.307-0.589
Social Life	0.348*	0.235-0.462
Travelling	0.277*	0.164-0.390

EQ-5D, EuroQol Five Dimension Questionnaire; VAS, Visual Analog Scale; NRS, Numeric Rating Scale; COMI, Core Outcome Measures Index; NDI, Neck Disability Index; ODI, Oswestry Disability Index; K_w, Weighted kappa; CI, Confidence Interval;*, $p<0.05$

All Patients (N=177)										
Collected Preoperative	0.288 [0.317]	50 [33]								
Recalled Preoperative	0.068* [0.402]	30* [30]								
Postoperative	0.482* [0.479]	70* [35]								
Cervical (N=46)										
Collected Preoperative	0.288 [0.239]	50 [33]	7.4 [3.1]	6 [6]	6 [6]	44 [26]				
Recalled Preoperative	0.293 [0.463]	40 [40]	7.9 [3.8]	7 [7]	7 [5]	48* [34]				
Postoperative	0.459* [0.409]	60* [30]	4.4* [4.7]	4.5* [6]	5* [8]	34* [35]				
Lumbar (N=131)										
Collected Preoperative	0.288 [0.366]	50 [34]					7.9 [2.4]	7 [3]	8 [3]	52 [29]
Recalled Preoperative	0.004* [0.363]	30* [30]					9.0* [1.4]	8* [2]	9* [2]	64* [22]
Postoperative	0.555* [0.712]	70* [35]					4.2* [5.7]	4* [7]	4* [7]	26* [41]
Neurogenic claudication (N=54)										
Collected Preoperative	0.288 [0.196]	50 [36]					8.1 [1.8]	7 [2]	8 [3]	54 [21]
Recalled Preoperative	0.007* [0.401]	30* [30]					8.8* [1.5]	8* [2]	9* [2]	63* [23]
Postoperative	0.446* [0.538]	70* [30]					5.1* [6.1]	5* [6]	5* [6]	34* [46]
Lumbar Radiculopathy (N=77)										
Collected Preoperative	0.287 [0.378]	50 [32]					7.9 [2.4]	7 [4]	8 [3]	50 [32]
Recalled Preoperative	-0.018* [0.446]	30* [30]					9.0* [1.3]	8* [2]	9* [2]	64* [24]
Postoperative	0.592* [0.675]	70* [40]					4.1* [5.6]	3* [7]	2* [7]	24* (41]

EQ-5D, EuroQol Five Dimension Questionnaire; VAS, Visual Analog Scale; NRS, Numeric Rating Scale; COMI, Core Outcome Measures Index; NDI, Neck Disability Index; ODI, Oswestry Disability Index; *, $p<0.05$; all p values are relative to the collected preoperative score; interquartile range is given in square brackets. Wilcoxon signed-rank test was used.

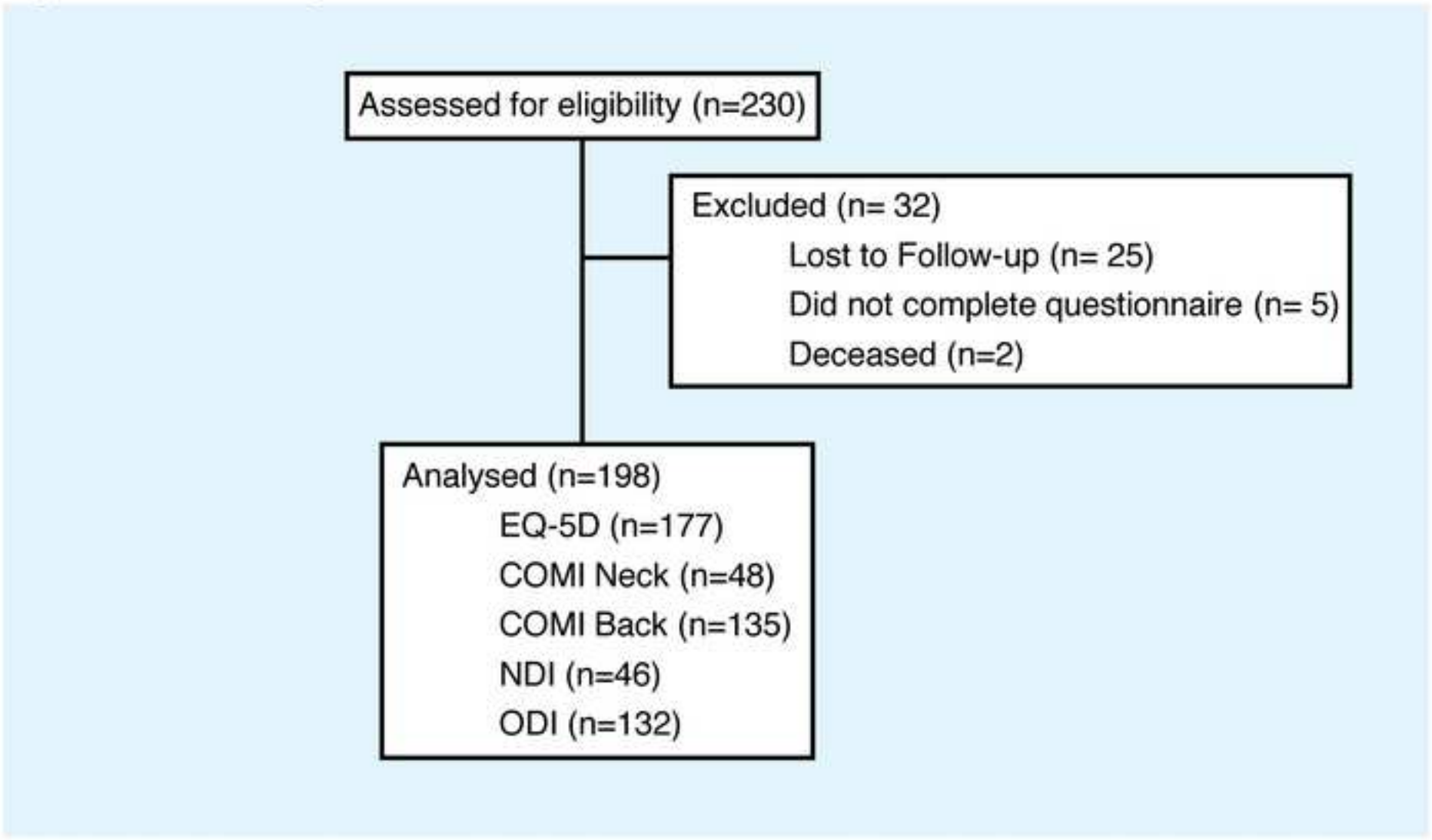
TABLE 5. Comparison of Recalled to Collected Preoperative Median Scores by Self-Reported Outcome of Surgery

	EQ-5D	Health VAS	COMI Neck	Neck Pain NRS	Arm/Shoulder Pain NRS	NDI	COMI Back	Back Pain NRS	Leg/Buttock Pain NRS	ODI
Helpful Surgery										
All Patients (N=155)										
Collected Preoperative	0.288 [0.317]	50 [34]								
Recalled Preoperative	0.007* [0.402]	30* [30]								
Cervical (N=38)										
Collected Preoperative	0.288 [0.314]	50 [34]	7.3 [3.1]	6 [5]	6.5 [5]	38 [21]				
Recalled Preoperative	0.288 [0.463]	45 [43]	8.0 [3.6]	7 [7]	7 [5]	52* [33]				
Lumbar (N=117)										
Collected Preoperative	0.288 [0.329]	50 [35]					7.9 [2.3]	7 [3]	8 [3]	51 [29]
Recalled Preoperative	-0.018* [0.300]	30* [30]					9* [1.4]	8* [3]	9* [2]	66* [22]
Neurogenic claudication (N=49)										
Collected Preoperative	0.288 [0.197]	50 [35]					8 [1.8]	7 [2]	8 [3]	54 [22]
Recalled Preoperative	0.007* [0.401]	30* [30]					8.9* [1.6]	8* [2]	8* [2]	66.5* [25]
Lumbar Radiculopathy (N=68)										
Collected Preoperative	0.288 [0.378]	50 [35]					7.7 [2.8]	7 [4]	7.5 [4]	46 [31]
Recalled Preoperative	-0.046* [0.433]	30* [30]					9* [1.4]	8* [3]	9* [2]	65* [27]
Unhelpful Surgery										
All Patients (N=22)										
Collected Preoperative	0.288 [0.402]	50 [38]								
Recalled Preoperative	0.288 [0.280]	40 [31]								
Cervical (N=8)										
Collected Preoperative	0.343 [0.185]	50 [30]	8.0 [3.8]	4.5 [7]	4 [7]	51 [40]				
Recalled Preoperative	0.325 [0.449]	40 [31]	7.5 [3.8]	7 [8]	7.5 [8]	40 [32]				
Lumbar (N=14)										
Collected Preoperative	0.226 [0.296]	45 [42]					8.8 [1.7]	7 [4]	9 [2]	62 [23]
Recalled Preoperative	0.229 [0.368]	35 [28]					8.6 [1.1]	8 [2]	9 [2]	58 (23]

EQ-5D, EuroQol Five Dimension Questionnaire; VAS, Visual Analog Scale; COMI, Core Outcome Measures Index; NDI, Neck Disability Index; ODI, Oswestry Disability Index; *, $p<0.05$; all *p* values are relative to the preoperative score; interquartile range is given in square brackets.

Figure 1
Click here to download high resolution image

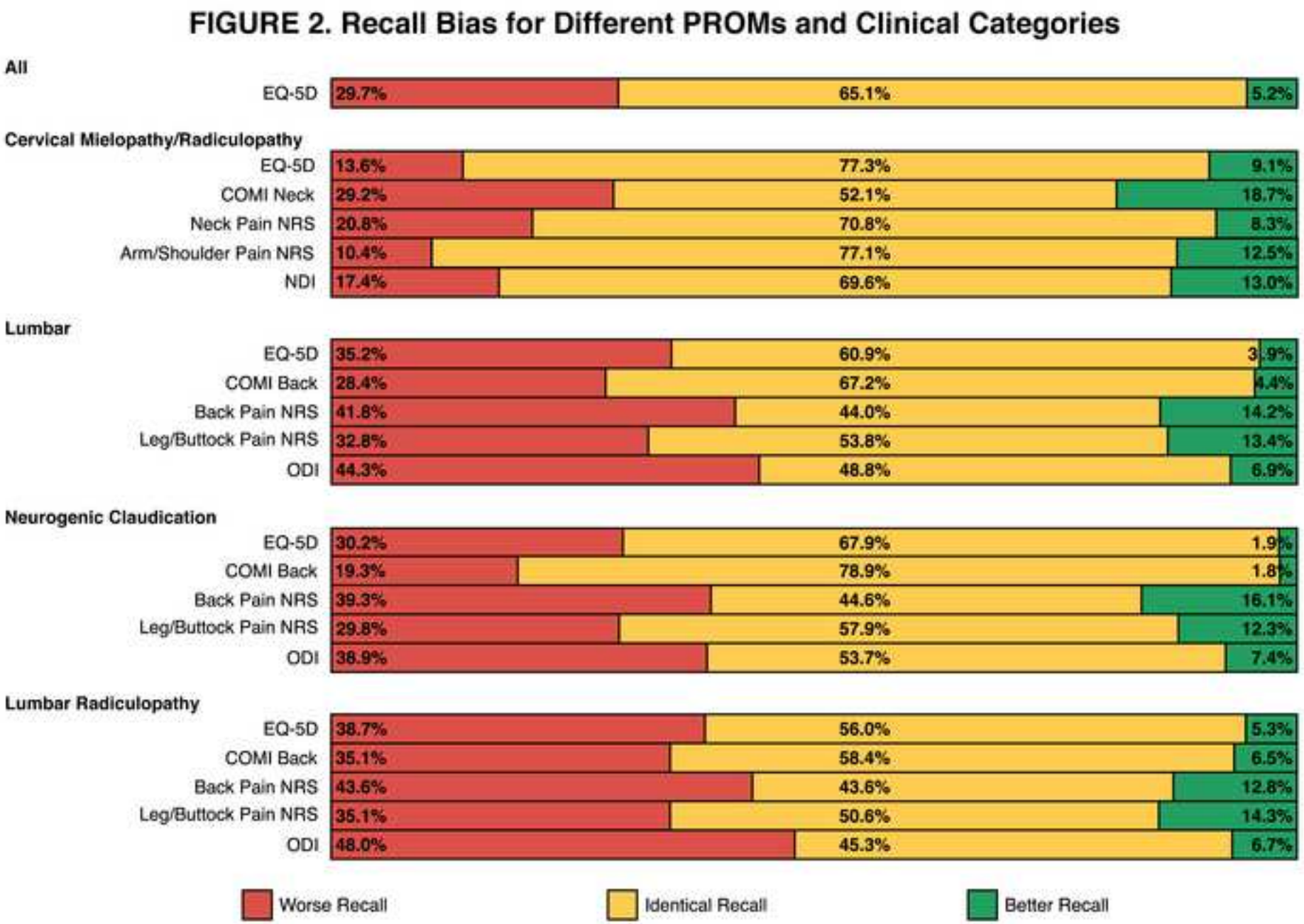
Figure 1 – Flow Diagram



EQ-5D, EuroQol Five Dimension Questionnaire; COMI, Core Outcome Measures Index; NDI, Neck Disability Index; ODI, Oswestry Disability Index.

Figure 2

Click here to download high resolution image



EQ-5D, EuroQol Five Dimension Questionnaire; VAS, Visual Analog Scale; NRS, Numeric Rating Scale; COMI, Core Outcome Measures Index; NDI, Neck Disability Index; ODI, Oswestry Disability Index. Recall bias was calculated acccording to the MCID value.